IN VITRO EXCISION OF ADENO-ASSOCIATED VIRUS DNA FROM RECOMBINANT PLASMIDS: ISOLATION OF AN ENZYME FRACTION FROM HELA CELLS THAT CLEAVES DNA AT POLYPURINE-POLYPYRIMIDINE SEQUENCES RICH IN GC BASE PAIRS

BY

H. JONATHAN GOTTLIEB

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1988

To my family

Digitized by the Internet Archive
in 2011 with funding from
University of Florida, George A. Smathers Libraries with support from LYRASIS and the Sloan Foundation

http://www.archive.org/details/invitroexcisiono00gott

ACKNOWLEDGEMENTS

I would like to express my sincerest appreciation to my mentor and friend, Dr. Nicholas Muzyczka, for the unique opportunity to work with him during the last several years. In addition, I would like to thank the members of my graduate committee at the University of Florida and Drs. Kenneth Berns and William Holloman for their helpful suggestions and encouragement.

I am grateful to members of the Muzyczka laboratory in New York and in Florida for their advice, conversation and friendship. Special thanks are extended to Deirdre Lawe and to my co-defendant, Susan McLaughlin.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

ATP	adenosine-5'-triphosphate
bp	base pairs
BSA	bovine serum albumin
CTP	cytosine-5'-triphosphate
d	daltons
dATP	deoxyadenosine-5'-triphosphate
dCTP	deoxycytosine-5'-triphosphate
DEAE	diethylaminoethyl
dGTP	deoxyguanosine-5'-triphosphate
DNA	deoxyribonucleic acid
DTT	dithiothreitol
dTTP	thymidine-5'-triphosphate
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetra-acetic acid
form I	superhelical plasmid DNA
form II	nicked circular DNA
form III	linear duplex DNA
G:C	double-stranded poly(dG):poly(dC)
GTP	guanosine-5'-triphosphate
kb	kilobase
kd	kilodaltons
PMSF	phenylmethylsulfonylfluoride

POLY(dA)	polydeoxyadenylic acid
POLY(dC)	polydeoxycytidylic acid
POLY(dG)	polydeoxyguanylic acid
POLY(dT)	polydeoxythymidylic acid
RNA	ribonucleic acid
rpm	revolutions per minute
SDS	sodium dodecyl sulfate
ssb	*E. coli* single-stranded binding protein
telomerase	telomere terminal transferase
Z-DNA	left-handed helical double-stranded DNA

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

IN VITRO EXCISION OF ADENO-ASSOCIATED VIRUS DNA FROM RECOMBINANT PLASMIDS: ISOLATION OF AN ENZYME FRACTION FROM HELA CELLS THAT CLEAVES DNA AT POLYPURINE-POLYPYRIMIDINE SEQUENCES RICH IN GC BASE PAIRS

By

H. Jonathan Gottlieb

August 1988

Chairman: Nicholas Muzyczka
Major Department: Immunology and Medical Microbiology

When circular recombinant plasmids containing adeno-associated virus (AAV) DNA sequences are transfected into human cells, the AAV provirus is rescued. Using these circular AAV plasmids as substrates, a cellular site-specific endonuclease, endo R, was isolated from HeLa cell nuclei on the basis of its ability to excise intact AAV sequences *in vitro* from the vector DNA and produce linear DNA products. The enzyme recognizes and cleaves polypurine-polypyrimidine sequences which are at least 9 residues long and rich in GC base pairs. Such sequences occur naturally in non-coding regions of higher eucaryotes, are present in AAV recombinant plasmids as part of the first 15 bp of the AAV terminal repeat and, in some cases, are present at the AAV/vector junction as the result of cloning by GC tailing. Plasmid DNA that is transfected into tissue culture cells is cleaved *in vivo* to produce a pattern of DNA fragments similar to that seen with purified enzyme *in vitro*. The level

of specific cleavage in the standard assay depends on the type and length of the recognition signal in the substrate. Generally, longer and more homogeneous stretches of poly(dG):poly(dC) are better substrates for double-stranded cleavage. The enzyme apparently recognizes the secondary structure characteristic of G-rich polypurine sequences in duplex DNA. Sequence analysis of the cut site suggests that double-stranded cleavage occurs through a series of concerted single-stranded nicks of the substrate throughout the enzyme recognition site.

The molecular weight of the active form of endo R was estimated by gel filtration and sedimentation in glycerol gradients to be approximately 125,000. The enzyme, which requires magnesium as a cofactor, does not have significant endonucleolytic activity on single-stranded DNA, but is equally active on closed circular and linear duplex DNA substrates. The properties of endo R are consistent with it playing a role in the rescue of integrated AAV sequences. The enzyme's function in cellular DNA metabolism is less obvious, but its properties suggest that it may have a role in cellular DNA replication and recombination.

CHAPTER I
INTRODUCTION

Introduction

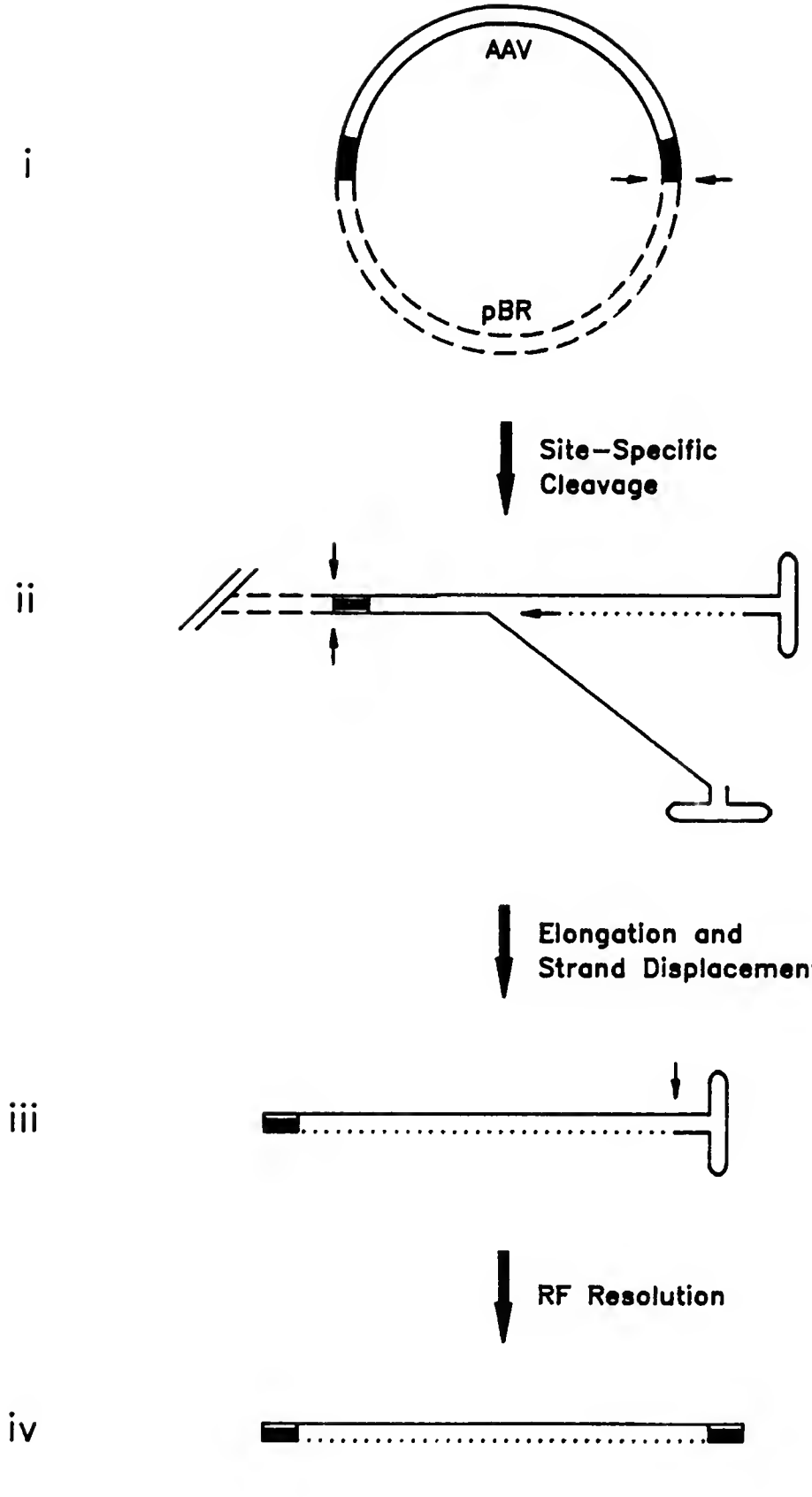
A major obstacle in the characterization of cellular factors involved in mammalian DNA metabolism lies in establishing a biochemically and genetically defined system. For the most part, the complexity of cellular processes and the lack of genetic analysis in higher eucaryotes prevent accurate characterization of these processes. This problem has been partially circumvented with the use of well characterized mammalian viral systems as probes for cellular functions (Challberg and Kelly, 1979; Li and Kelly, 1984; Murakami et al., 1986). In this regard, adeno-associated virus (AAV) is ideal. AAV has been well characterized genetically (Hermonat et al., 1984), and because of its small size, is largely dependent on cellular functions for viral multiplication (Berns et al., 1985).

AAV DNA Replication

The current model for AAV DNA replication is a modified version of a general model for the replication of eucaryotic DNA molecules first proposed by Cavalier-Smith (Cavalier-Smith, 1974; Hauswirth and Berns, 1977; Straus et al., 1976). AAV contains a 4.7 kb single-stranded linear genome consisting of an internal nonrepetitive sequence which is flanked at each end by inverted terminal repeats that are palindromic (Lusby et al., 1980). DNA replication proceeds by leading strand

Figure 1-1. Model for AAV DNA Replication.

The model for AAV DNA replication is a modified version of the model for eukaryotic DNA replication first described by Cavalier-Smith (1974). i) Single or double-stranded specific cleavage at the AAV/pBR322 junction is necessary to allow formation of the hairpin primer in the palindromic region of the termini. Site specific cleavage must occur at the other AAV/vector junction either before or during DNA synthesis to separate AAV and vector sequences. ii) DNA replication proceeds from the 3′ end of the primer to the other end of the molecule and includes synthesis of the distal terminal repeat. iii) A specific nick is required on the parental strand directly across from the original 3′ end of the primer to allow repair of the parental terminal sequences (iv). Both the parental and progeny strands are now capable of initiating another DNA replication cycle.


i
AAV
pBR
Site–Specific
Cleavage
ii
Elongation and
Strand Displacement
iii
RF Resolution
iv

synthesis from the hairpin primer formed in the palindromic region of the terminal repeat (Figure 1-1). Strand elongation proceeds from the 3′ end of the primer, in a 5′ to 3′ direction, to the other end of the molecule. A specific nick is required in the parental strand at a point directly across from the original 3′ terminal base to allow repair of the parental terminal sequences and to produce a full-length duplex replicative intermediate. Subsequently, both the parental and progeny strands are capable of initiating a second round of DNA replication and displacing a progeny single-stranded DNA molecule.

Rescue of Latent AAV Provirus

AAV is a defective parvovirus and requires the presence of a co-infecting helper virus for a productive viral infection. Virtually any member of the adenovirus or herpes virus family can supply the helper functions (Atchison et al., 1965; Buller et al., 1981; McPherson et al., 1985). In the absence of a helper virus, AAV readily integrates into the host DNA (Hoggan et al., 1972; Berns et al., 1975; Handa et al., 1977), either as a single proviral copy or more often as a tandem head to tail array of several AAV genomes (Cheung et al., 1980; Laughlin et al., 1986; McLaughlin et al., 1988). Super-infection of latently infected cells with helper virus results in the rescue of integrated provirus and normal AAV replication (Hoggan et al., 1972).

In many proviral cell lines the yield of replicative intermediates and virus produced by superinfection with helper virus are identical to those obtained from exogenous AAV infections (Laughlin et al., 1986; McLaughlin et al., 1988). This suggested that the rescue of AAV sequences from chromatin is a rapid event initiated by specific cleavage

within the AAV termini. Additional evidence that AAV rescue involves a site-specific nuclease comes from the study of recombinant AAV plasmids which, presumably, rescue by a mechanism similar to that used by proviruses integrated into chromatin (Samulski et al., 1982). When AAV-pBR322 plasmids are transfected into human cells in the presence of adenovirus, free linear duplex AAV DNA, the major replicative form (Straus et al., 1976; Hauswirth and Berns, 1977), is seen within 24 hours (Samulski et al., 1982). However, when viable terminal deletion mutants are transfected into adenovirus infected cells, the formation of replicative forms (RF) is delayed (Samulski et al., 1983; Lefebvre et al., 1984, Samulski et al., 1987), indicating that either gene conversion of the terminal deletion causes a delay in DNA synthesis, or that AAV sequences at the termini are required for excision. Studies with the AAV mutant, psub201^{+}, in which the terminal 13 bp have been deleted from both ends, suggest that the delay in RF formation is due to a defect in rescue (Samulski et al., 1987) and that the signal for AAV excision resides in the terminal sequences. AAV DNA replication in adenovirus infected cells transfected with psub201^{+} was eight-fold more efficient when the AAV sequences were cleaved from the vector sequences prior to transfection (Samulski et al., 1987). Presumably, cleavage of the plasmid bypassed the excision step which must normally precede DNA replication.

In principle, the duplex replicative intermediate can be generated from the covalently closed form I plasmid DNA in one of two ways. Either the AAV sequences are separated from the plasmid by precise excision or a single-stranded DNA molecule is generated by AAV-specific replication. (See Samulski et al., 1983, and Senepathy et al., 1984,

for examples of the second mechanism.) Because the AAV termini are also the origins for AAV DNA replication (Hauswirth and Berns, 1977; Samulski et al., 1983; Senepathy et al., 1984), it is difficult to distinguish between the two mechanisms. Both mechanisms, however, require either a specific nick or a double-stranded cut at an AAV/vector junction in the input plasmid DNA (Figure 1-1). Thus, it should be possible to use AAV recombinant plasmids as substrates for an in vitro excision assay and to screen cellular extracts for an activity that produces either single-stranded nicks or double-stranded cuts at the AAV/vector junctions.

General Aim of the Studies

Using the recombinant AAV plasmids as substrates, an enzyme fraction was isolated from HeLa cell nuclear extracts that rescued intact AAV DNA from vector DNA in vitro and produced linear DNA products. The double-stranded cleavage activity has been called endo R because of its ability to rescue AAV sequences. The activity was of cellular origin and was stimulated approximately 5 fold by Ad2 infection in the presence of hydroxyurea.

In the following chapters, the purification and properties of the enzyme are discussed and the activity observed in vitro is compared with what is observed in vivo. In addition, the behavior of the enzyme suggests that it may be involved in cellular DNA metabolism, and these properties are discussed as well.

CHAPTER II
MATERIALS AND METHODS

Cell Culture and Viruses

HeLa S3 cells were maintained at 37°C in suspension culture in Eagles minimal essential medium supplemented with 5% calf serum, 1% glutamine, penicillin and streptomycin. Wild type adenovirus 2 (Ad2) was prepared from a freeze/thaw lysate of HeLa S3 cells as previously described (Samulski et al., 1983). HeLa monolayer cells were transfected with 5.0 μg DNA by the DEAE-dextran method (McCutchan and Pagano, 1968) as described (Muzyczka, 1980), and infected with Ad2 virus at a multiplicity of infection (moi) of 10. Low molecular DNA was isolated from the cells by the method of Hirt (1967) as described (Muzyczka, 1980).

Recombinant Clones and Enzyme Substrates

All recombinant clones were maintained in either the recA host HB101 (Boyer, 1969) or the recBC, recF, sbcB host JC8111 (Boissy and Astell, 1985), a gift from Peter Tattersall. This was to prevent variation in the size of the G:C tails and other repetitive inserts (Hauswirth et al., 1984) and to prevent deletion of the palindromic AAV terminal repeat (Samulski et al., 1982).

The AAV wild-type plasmid pSM620 contains a full copy of the AAV genome inserted by GC tailing into the PstI site of pBR322 (Bolivar et al., 1979). The plasmid contains 18 bp of poly(dG):poly(dC) (G:C) at

the left pBR322/AAV junction and 28 bp of G:C at the right junction (Samulski et al., 1983). The AAV terminal mutant clone, pSM703, was isolated by Samulski et al. (1983) and was sequenced as part of this study. It contains 100 bp deletions in both AAV terminal sequences, 38 bp of G:C tail at the right AAV/pBR junction, and 36 bp G:C at the left junction. The plasmids pGM620C and pGM620D contain either the left or the right terminal PstI fragment, respectively, of the wild-type AAV plasmid, pSM620 (Samulski et al., 1983), subcloned into the PstI site of pBR322 (Bolivar et al., 1979). Both clones contain the original poly(dG)-poly(dC) tails that are present in pSM620. The plasmid pGM1008 was constructed by removing the PstI-BssHII fragment (AAV nucleotides 4254 to 4657) from pGM620D and religating the molecule after producing blunt ends with the Klenow fragment of DNA polymerase I. The plasmid contains the parental poly(dG)-poly(dC) tail plus the terminal 21 bp of the AAV sequence up to the BssHII site at nucleotide 4657.

The plasmids pGM913, pGM1116, pGM1505, pGM1635, pGM1228, pGM1344 and pGM1483 were constructed by inserting chemically synthesized oligonucleotides (Systec) into a unique restriction site in pBR322. To construct these clones, 2 pmol of single-stranded oligomer was boiled with an equal amount of the appropriate complementary strand in 20 μl of water for 5 minutes. The reaction was then cooled slowly to room temperature to allow the formation of duplex molecules. Ligation reaction mixtures containing 2 pmol of annealed oligomer, 0.2 pmol of PstI linearized pBR322 and 400 units of T4 DNA Ligase (New England BioLabs) were incubated at 15^{o}C for 12-24 hours. The 5' ends of the single-stranded oligomers were phosphorylated only when multiple tandem copies of the insert were desired. Dephosphorylation of the vector was

generally not required under these conditions. Positive clones retained tetracycline resistance, but were sensitive to ampicillin.

The plasmids PGM913 and pGM1116 contain 13 and 9 base pairs of G:C homopolymer, respectively. PGM1635 contains 20 bp of alternating copolymer GC, and pGM1505 contains the Dictyostelium telomeric repeat sequence $(C_{2-6}T)_5$ (Shampay et al., 1984). Two plasmids contain the sequence of the terminal 21 bp of AAV (ggCCaCTCCCTCTCTgCgCgC), in either a monomer form (pGM1228) or as an inverted dimer (pGM1344). All of the above clones were constructed by inserting the oligomers into the PstI site of pBR322. Additionally, the plasmid pGM1483 contains an insert of the Tetrahymena telomeric sequence, $(C_4A_2)_3$ (Blackburn and Szostak, 1984), chemically synthesized and inserted into the EagI site of pBR322. In each case, both strands of the oligonucleotides were synthesized so that the restriction enzyme recognition site was maintained after the two strands of each oligonucleotide were annealed and ligated into pBR322. The inserts and pBR322 flanking sequences of these plasmids were confirmed by DNA sequencing.

The plasmid pGA38, a gift from Todd Evans, contains an insert of the alternating co-polymer $(GA)_{38}$, cloned into the EcoRI site of pUC9 (Evans and Efstratiatis, 1986). PGA11 is a subclone of pGA38 and contains an insert of $(GA)_{11}$ cloned into the EcoRI site of pUC9. The plasmid pEV136, a gift from Eckhardt Wimmer, is an infectious polio clone and contains an 18 bp G:C insert at the EcoRI site, as well as an 84 bp stretch of poly(dA):poly(dT) (Semler et al., 1984).

Purification of Endo R

All operations were carried out at 0-4°C. To minimize the effects of proteolysis, all buffers contained 0.1 mM phenylmethylsulfonyl-fluoride (PMSF) and the time between chromatographic steps was kept to a minimum. Nuclei from eight liters of adenovirus infected HeLa S3 cells, grown in the presence of 10 mM hydroxyurea (Sigma Chemical Co.), were isolated as described by Challberg and Kelly (1979). Frozen nuclei were thawed on ice and 5 M NaCl was added to a final concentration of 0.2 M. After incubation on ice for 2 hours, the nuclei were pelleted by centrifugation at 3000 x g for 15 minutes, the pellet was discarded and the supernatant was dialyzed against 25 mM Tris-HCl pH 8.0, 10 mM KCl, 0.1 mM EDTA, 5 mM 2-mercaptoethanol, 0.1 mM PMSF and 20% glycerol (Buffer A). Any insoluble precipitate in the nuclear extract was removed by centrifugation at 12000 x g for 20 minutes and the supernatant was retained as fraction I. Fraction I contained 34.2 mg/ml protein, as determined by the Bradford protein concentration assay (Bradford, 1976), in a volume of 25 ml.

Fraction I was loaded onto a 15 ml DEAE cellulose column previously equilibrated with buffer A, pH 8.0. The column was then washed with 3 column volumes of buffer A and eluted in the same buffer with a 150 ml linear gradient containing 0.01 to 0.5 M KCl. Fractions emerging between 0.1 and 0.25 M KCl were pooled to form fraction II and contained 2.7 mg/ml protein a total volume of 60 mls. Fraction II was dialyzed against two 1-liter changes of buffer B (10 mM sodium phosphate, pH 6.5, 5 mM 2-mercaptoethanol, 10% glycerol and 0.1 mM PMSF) containing 0.01 M NaCl and loaded onto a 7 ml phosphocellulose column pre-washed with 21 mls of the same buffer. The column was eluted with a

70 ml linear gradient from 0.01 to 1 M NaCl in buffer B. Active fractions emerged about halfway through the gradient and were pooled to form fraction III. Fraction III contained 0.16 mg/ml protein in 45 ml and had a specific endo R activity of approximately 3000 units/mg. Fraction III was split into 2 aliquots and each aliquot was concentrated about 55 fold using an Amicon Centriflo 25 spin concentrator, yielding a final protein concentration of 18 mg/ml in 2.0 ml. The two 1.0 ml aliquots of concentrated fraction III were passed separately through a 36 ml (1.5 x 45 cm) column of Sephadex G200 equilibrated with buffer B, pH 6.5, 250 mM NaCl. Fractions containing endo R activity were pooled (fraction IV), dialyzed against buffer B containing 0.01 M NaCl and loaded onto a 3 ml poly(dG) agarose column equilibrated in buffer B, 10 mM NaCl. The column was washed with 3 bed volumes of the same buffer and eluted with a 30 ml linear gradient from 0.01-1.0 M NaCl. Active fractions were pooled (fraction V) and dialyzed into 25 mM Tris-HCl, pH 7.5, 50% glycerol, 0.1 mM DTT. Fraction V was stable for at least 2 months at -20°C without measurable loss of activity.

Nuclease Assays

The standard endo R reaction mixtures of 25 μl contained 20 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 1 mM dithiothreitol, 0.2 pmol form I plasmid substrate and 0.1-5.0 units of endo R (see below). After incubation at 37°C, the reaction products were phenol extracted, ethanol precipitated and digested with either BstEII at 60°C for 1 hour or with SphI at 37°C for 3 hours. The reaction products were then fractionated by electrophoresis on 1.4% agarose gels. One unit of endo R activity is defined as that amount of protein that is required to cleave 50% of

pGM620D substrate under standard conditions. During early stages of the purification, it was necessary to treat the reaction products with 0.25 mg/ml proteinase K (Sigma) for 1 hour at 37°C prior to phenol extraction and gel electrophoresis. When indicated, endo R cleavage reactions were terminated with the addition of an equal volume of a 2x proteinase K stop solution, containing 0.5 mg/ml Proteinase K, 1% SDS and 20 mM EDTA. After addition of the stop solution, incubation was continued at 37°C for at least 1 hour.

Single-stranded nuclease assays contained 5.0 μg of heat-denatured ^{3}H-labeled E. coli chromosomal DNA (1 x 10^{5} cpm/μg) in 250 μl. For S1 nuclease, the mixtures contained 50 mM sodium acetate (pH 4.5), 0.3 M NaCl, 10 mM $ZnCl_2$, and 1.25 units of S1, while the endo R reaction was under standard reaction conditions. Twenty-five microliter portions were removed at the indicated times and added to 2ml of 10% trichloroacetic acid (TCA) containing 200 μg/ml BSA. The precipitates were collected on nitrocellulose filters (Schleicher and Schuell type BA85), washed with TCA and ethanol and then counted. E.coli DNA was labeled with ^{3}H-thymidine (6.7 Ci/mmole, ICN) and isolated as described (Holloman et al., 1981).

DNA Labeling, Sequencing and Primer Extension

Nucleotide sequences were determined by the method of Maxam and Gilbert (1977) and by the dideoxy method (Sanger et al., 1977). For Maxam and Gilbert sequencing of the pGM clones, 10 μg of DNA was linearized with ScaI, dephosphorylated with calf intestine alkaline phosphatase (CIAP) and labeled at the 5' end with polynucleotide kinase and γ^{32}P-ATP or labeled at the 3' end with the Klenow fragment, α^{32}P-

dCTP and 80 μM each of dGTP, dATP and TTP. Five micrograms of the labeled DNA was further digested with BamHI and the 5′ or 3′ labeled ScaI-BamHI fragments were isolated from 1.2% low melting agarose prior to sequencing. Endo R cleavage fragments were prepared by incubating the remaining 5 μg of 5′ or 3′ labeled DNA with 10 units of fraction V endo R under standard reaction conditions and isolating the labeled endo R/ScaI fragments from a 4% non-denaturing acrylamide gel as described (Maniatis et al., 1982). The sequence at the site of cleavage was determined directly from the sequencing gels by comparing the mobility of the endo R fragments and the sequence markers.

For primer extension of endo R products, 2 μg of plasmid DNA was incubated with endo R in a standard reaction mixture of 50 μl. The form III product was then isolated from 1.2% low melting agarose and dissolved in 10 μl of water and 4 μl of 5x reaction buffer (0.3 M Tris-HCl, pH 8.3; 0.375 M NaCl; 37.5 mM $MgCl_2$ and 2.5 mM DTT). The reaction mixtures were divided into two 7 μl portions and 1 μl of either upper or lower strand pBR322/PstI primers (2 pmol/μl) was added to each portion. The solutions were heated to 100°C for 5 minutes under parafin oil and immediately frozen in a dry ice/ethanol bath. After thawing on ice, 1 μl each of stock solutions containing 0.5 mM dGTP, 0.5 mM dCTP, 0.5 mM TTP, 0.02 mM dATP; 10 μCi/μl α^{32}P-dATP (>3000 Ci/mM) and 10 units/μl AMV reverse transcriptase (IBI) were added to each portion to produce a final reaction volume of 10 μl. After incubation at 42°C for 10 minutes, 2 μl of chase solution (0.25 mM each of dGTP, dATP, dCTP and TTP) was added and incubation was continued at 42°C for an additional 10 minutes. The reaction was terminated with the addition of 7 μl of stop buffer containing 1.6 μl 0.25 M EDTA; 2 μl 3 M sodium acetate; 0.2 μl 2

mg/ml tRNA and 3.2 μl water. After ethanol precipitation with 3 volumes of 95% ethanol, the primer extension products were redissolved in 5 μl of sequence gel running buffer (Maxam and Gilbert, 1977) and heated to 100°C for 3 minutes prior to gel electrophoresis. The sequence at the site of endo R cleavage was determined by comparing the mobility of the primer extension products with dideoxy sequencing ladders produced using identical single-stranded primers.

Gel Electrophoresis

Electrophoresis in the presence of sodium dodecyl sulfate was done according to Laemmli (1970) in slab gels of 7.5% acrylamide, 0.2% N,N'-methylene-bisacrylamide. The gels were fixed in 50% methanol overnight and silver stained as described (Switzer et al., 1979).

Neutral polyacrylamide and agarose gel electrophoresis of DNA was as described (Muzyczka, 1980). DNA fractionated on agarose gels were transferred to nitrocellulose filters (Southern, 1975) and hybridized to ^{32}P-nick translated DNA probe (>2x10^8 cpm/μg) as described (Muzyczka, 1980).

Denaturing polyacrylamide-urea gels were as described (Maxam and Gilbert, 1977). All sequencing gels were run at 55-60°C to prevent compression of palindromic sequences (Lusby et al., 1980).

For denaturing agarose gels, 1.5% horizontal agarose gels were poured in 50mM NaCl and allowed to equilibrate in excess running buffer (30 mM NaOH, 1 mM EDTA) for at least 1 hour. Samples were incubated for 5 minutes in gel running buffer (Maniatis et al., 1982) containing 0.2 N NaOH prior to electrophoresis at 30 volts for at least 24 hours.

Estimation of the Isoelectric Point

Estimation of the pI of endo R was performed by a modified version of the method by Yang and Langer (1987). Eight 100 μl aliquots of packed DEAE-cellulose (Whatman DE32) was equilibrated in 0.1 M Tris-HCl at pH 7.0, 7.5, 8.0 or 8.5 or in 0.1 M 2, 2'-Bis(hydroxymethyl)-2,2',2''-nitrilo triethanol (Bis-Tris) at pH 5.5, 6.0, 6.5 or 7.0. This required several changes of buffer until no change in the pH of the supernatant was observed. After equilibration, the cellulose was packed by centrifugation at 1000 x g for 1 minute and the supernatant was decanted. Mixtures of 20 μl of fraction III endo R and 80 μl of 0.1 M Tris or 0.1 M Bis-Tris buffer, at the appropriate pH, were added to the packed resin and the samples were incubated on ice for 30 minutes with frequent mixing to allow the protein to adsorb to the resin. After pelleting at 1000 x g for 1 minute, the supernatant was removed and assayed for endo R double-stranded cleavage activity as described above. The pI was determined to be equal to the lowest pH at which all the activity had been adsorbed from the supernatant.

Chromatographic Media, Enzymes and Materials

DEAE-cellulose (DE32) and phosphocellulose (P-11) were purchased from Whatman. Sephadex G200, poly(dG) agarose, and calf intestine alkaline phosphatase (CIAP) were from Sigma and Bio-Gel A1.5M was purchased from BioRad. All restriction endonucleases, T4 DNA ligase and polynucleotide kinase were purchased from New England Biolabs. The labeled nucleotides, α^{32}P-dNTPs and γ^{32}P-ATP (>3000 Ci/mmol) were from New England Nuclear and ^{3}H-thymidine (6.7 Ci/mmol) was from ICN.

Other Methods

Protein concentrations were determined by the procedure of Bradford (1976). The Biogel A1.5M (45.5 x 0.7 cm) and the Sephadex G200 (45.5 x 1.0 cm) columns were calibrated in 10 mM NaPO4, pH6.5; 0.25 M NaCl; 10% glycerol and 5 mM 2-mercaptoethanol. The excluded volume was determined with the use of Blue Dextran 2000 (molecular weight >2 million, Sigma) and the column was calibrated with the use of B-amylase (200 kd), alcohol dehydrogenase (150 kd), BSA (66 kd), carbonic anhydrase (29 kd) and cytochrome C (12.4 kd) as molecular weight standards (Sigma Chem. Co.).

Estimate of In Vivo Excision Frequency

A typical transfection uses 0.1-10 μg of plasmid DNA per 10 cm dish or approximately $3x10^9$-$3x10^{11}$ molecules per dish. Assuming 10^6 competent cells per dish, each competent cell will have $3x10^3$-$3x10^5$ molecules of plasmid DNA. This agrees reasonably well with the amount of plasmid DNA recovered from transfected cells as measured by filter hybridization. If the <u>in vitro</u> frequency of cleavage by endo R at an outboard site of AAV is 1%, cleavage at both outboard sites to produce an intact replicative form would be 0.01%. This would produce 0.3 to 30 molecules of excised AAV per transfected cell.

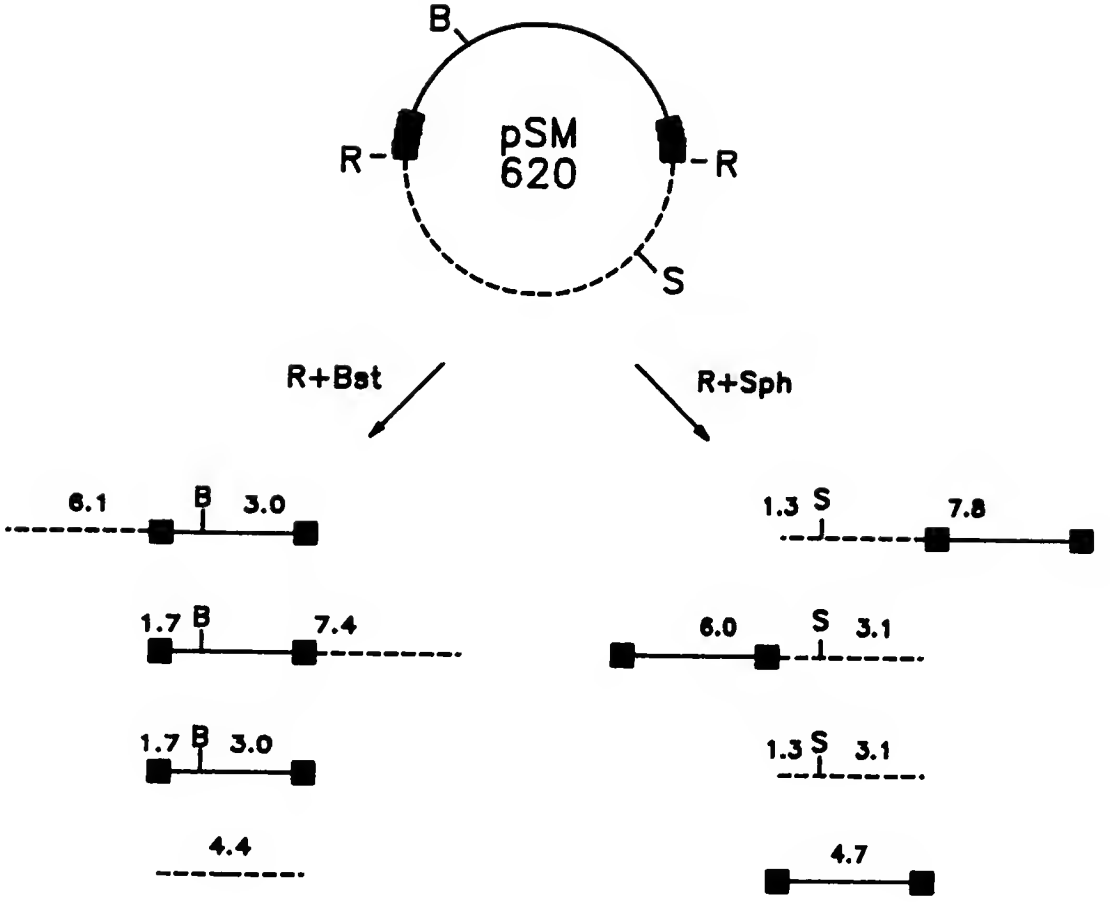
CHAPTER III
PURIFICATION OF ENDO R

Detection of a Site-Specific Endonuclease Activity in Crude Nuclear Extracts

The results from studies of AAV replication *in vivo* from latently infected cells and transfected AAV plasmids suggest that rescue from chromatin and from recombinant plasmids occurs by a similar mechanism (Samulski *et al.*, 1982) and that the sequences at or near the AAV termini are essential for this excision activity (Samulski *et al.*, 1987). Taking these observations into account, the wild type AAV plasmid, pSM620, was used as a substrate to assay for an activity that would specifically excise AAV sequences from the plasmid. The plasmid pSM620 contains the entire AAV genome cloned into the *Pst*I site of pBR322 by GC tailing (Samulski *et al.*, 1982). To detect site-specific cleavage at the AAV/vector junction, form I supercoiled DNA was treated with cell-free extract and digested with a one-cut restriction enzyme that cleaves either in the AAV sequences (*Bst*EII) or in vector DNA (*Sph*I). The products of the assay were then fractionated by agarose gel electrophoresis. Figure 3-1 illustrates the fragments that would be produced if cleavage occurs at either the left or right AAV/pBR junction. For example, cleavage with cellular extract at the right junction would generate a linear plasmid DNA molecule which, after *Bst*EII digestion, would produce a 3.0 kb fragment which consists exclusively of AAV DNA and a reciprocal 6.1 kb band that contains both AAV and pBR sequences.

Figure 3-1. Endo R Agarose Gel Assay.

The figure illustrates the products that would result from partial digestion by endo R and complete digestion by BstEII (left pathway) or SphI (right pathway) when the wild type plasmid, pSM620, is used as substrate. The thick line represents AAV DNA; the thin line represents pBR322 DNA; vertical lines represent cleavage sites of BstEII (B), SphI (S), and endo R (R). The filled squares represent the AAV terminal sequences. AAV cleavage occurs at the junction between the AAV termini and the vector DNA.


B
pSM
620
R
R
S
R+Bst
R+Sph
6.1
B
3.0
1.7
B
7.4
1.7
B
3.0
4.4
1.3
S
7.8
6.0
S
3.1
1.3
S
3.1
4.7

Using the agarose gel assay, crude extracts prepared from Ad2 infected, Ad2/AAV co-infected and uninfected HeLa cell nuclei were screened for the ability to cut near the AAV/vector junctions and the products were visualized by transfer to nitrocellulose filters and hybridization to AAV specific probe (Southern, 1975). Specific double-stranded cleavage activity was observed in uninfected and Ad/AAV co-infected cells at comparable levels. However, a 5-fold stimulation was observed in extracts from Ad2 infected cells treated with hydroxyurea. Whole cell extracts and extracts prepared from a 0.2 M NaCl wash of freeze-fractured nuclei (Challberg and Kelly, 1979) contained approximately equal amounts of activity, while nuclear extracts usually contained fewer contaminating nucleases. For these reasons, nuclear extracts from Ad2-infected HeLa cells treated with hydroxyurea were used as the primary source of activity.

Purification of Endo R

In the procedure outlined in Table I, crude nuclear extracts (fraction I) were prepared from Ad2-infected HeLa S3 cells, grown in the presence of 10mM hydroxyurea, and harvested at 21 hours post-infection (Challberg and Kelly, 1979). Due to the high level of contaminating endonucleases and cellular DNA (Figure 3-2), specific double-stranded cleavage activity could be detected at this stage only by Southern hybridization (Southern, 1975). Fraction I was applied to a column of DEAE cellulose and the activity that was eluted from the column was essentially free of nucleic acid and separated from at least one nonspecific cellular nuclease which did not bind to the column. Fraction II endo R activity (and the activity in subsequent fractions)

TABLE I

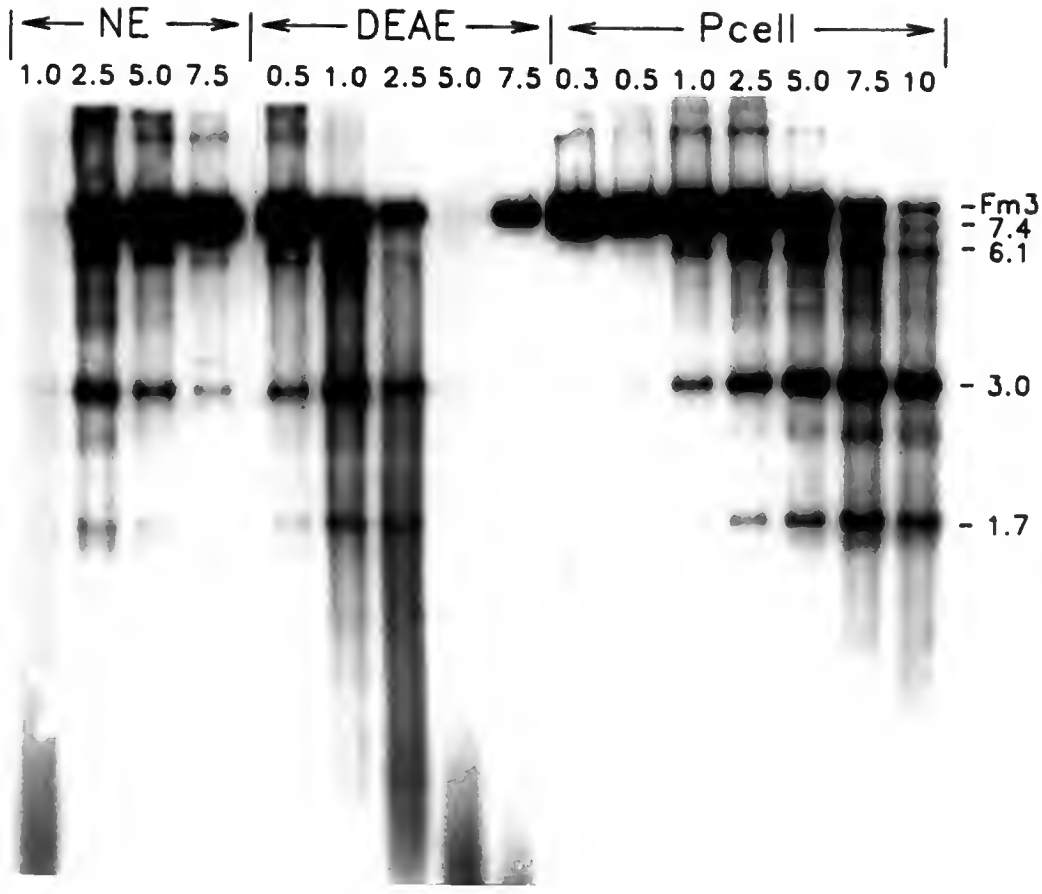
PURIFICATION OF ENDO R

	Fraction	Volume	Activity	Protein[a]	Specific Activity
		ml	units	mg	units/mg
I.	Nuclear Extract	25.0	60000	855.0	70
II.	DEAE-cellulose	60.0	60000	159.6	376
III.	Phosphocellulose	47.5	48000	38.5	1247
IV.	Sephadex G200	19.8	10000	9.1	1099
V.	Poly(dG) agarose	3.6	9000	2.9	3103

[a]Protein concentrations were determined by the method of Bradford (1976).

Figure 3-2. Enzyme Titration of Fractions I, II and III.

Standard reaction mixtures of 25 μl contained 0.1 pmol of pSM620 form I plasmid and the indicated amounts (μl) of nuclear extract (NE, Fraction I), DEAE-cellulose fractions (DEAE, Fraction II) or phosphocellulose fractions (Pcell, Fraction III). Reactions were terminated at 1 hour with the addition of proteinase K stop solution (see Chapter II), phenol extracted, ethanol precipitated and digested with BstEII. The samples were fractionated on a 1.4% agarose gel, transferred to a nitrocellulose filter and hybridized to AAV-specific probe.


NE
DEAE
Pcell
1.0 2.5 5.0 7.5
0.5 1.0 2.5 5.0 7.5
0.3 0.5 1.0 2.5 5.0 7.5 10
Fm3
7.4
6.1
3.0
1.7

could be detected directly by ethidium bromide staining of the gel assay (Figure 3-2). Active fractions from DEAE cellulose chromatography usually contained an overlapping peak of nonspecific nucleolytic activity. Further fractionation on a column of phosphocellulose produced an activity (fraction III) that exhibited substantially more specific cleavage with little evidence of contaminating nucleolytic activity at low enzyme concentrations (Figure 3-2). Nevertheless, high levels of fraction III enzyme resulted in non-specific degradation of the DNA substrate (data not shown). In order to separate the remaining contaminating nucleolytic activities from endo R, fraction III was further purified by chromatography on Sephadex G200 (fraction IV) and poly(dG) agarose (fraction V). Endo R activity emerged from the Sephadex G200 column just behind the void volume. This step resulted in a substantial loss of total endo R activity, but has been retained as part of the purification protocol for qualitative purposes. Endo R emerges from the final column, poly(dG) agarose, as a single symmetrical peak, when eluted with a linear increasing salt gradient, with little additional loss of activity.

Fraction IV endo R is stable for at least 2 months at -20°C without measurable loss of activity. The stability of the enzyme is greatly reduced when the concentration of protein falls below 200 μg/ml. For this reason, fraction V enzyme is dialyzed into buffer containing 50% glycerol and supplemented with 200 μg/ml enzyme grade BSA for periods of storage of up to 2 months at -20°C.

The Endo R Agarose Gel Assay

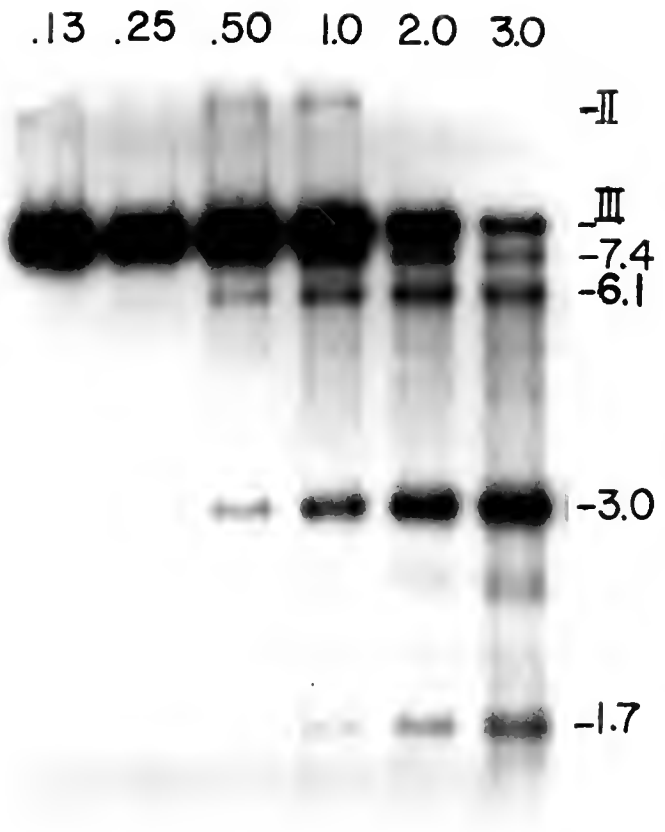
For the purposes of the gel assay, endo R can be thought of as a restriction enzyme with two sites in pSM620 (Figure 3-1). The products of endo R digestion can be seen as the products of a partial digestion of the substrate at either of the AAV/vector junctions. Further digestion of endo R products with the restriction enzyme BstEII, which cuts once asymmetrically in the AAV sequences, produces 6.1 and 3.0 kb fragments from endo R cleavage at the right AAV/pBR322 junction and 7.4 and 1.7 kb fragments from cleavage at the left junction. The formation of endo R products, using the plasmid pSM620 as substrate, is illustrated in figure 3-3 with the titration of partially purified enzyme (fraction III). As the reaction nears completion, most of the AAV/vector junctions are cleaved and produce the fragments expected from cleavage by BstEII and endo R. Linear fragments (form III) are the result of BstEII digestion of unreacted substrate. In further discussion, the production of the 3.0 and 1.7 kb fragments, which are easily monitored, is used to quantitate the amount of cleavage at the right and left junctions, respectively.

Molecular Size and Purity of Endo R

Estimates of the molecular weight of the native protein were made by chromatography of endo R on gel filtration columns and by centrifugation in glycerol gradients. The peak of endo R activity emerges from a Bio-Gel A1.5M column (see Chapter II: Materials and Methods) behind the 150 kdal alcohol dehydrogenase molecular weight marker with an apparent molecular weight of 120-125 kdal (data not shown). This is in good agreement with the observed behavior in a

Figure 3-3. Enzyme Titration of Fraction III.

The numbers at the top refer to the number of units of endo R used in a each reaction (see Chapter II: Materials and Methods). The sizes of the major cleavage fragments are indicated on the right. Standard endo R reaction mixtures (25 μl) contained 0.5 μg of form I pSM620 plasmid DNA, 5 mM $MgCl_2$, 20 mM Tris-HCl (pH 7.5), 1 mM DTT, and the indicated amount of fraction III enzyme. After incubation for 1 hour at 37^{o}C, the reaction products were treated with phenol, precipitated with ethanol, and digested with BstEII. The reaction mixture was then fractionated on a 1.4% agarose gel, transferred to a nitrocellulose filter and hybridized to nick-translated ^{32}P-AAV DNA.

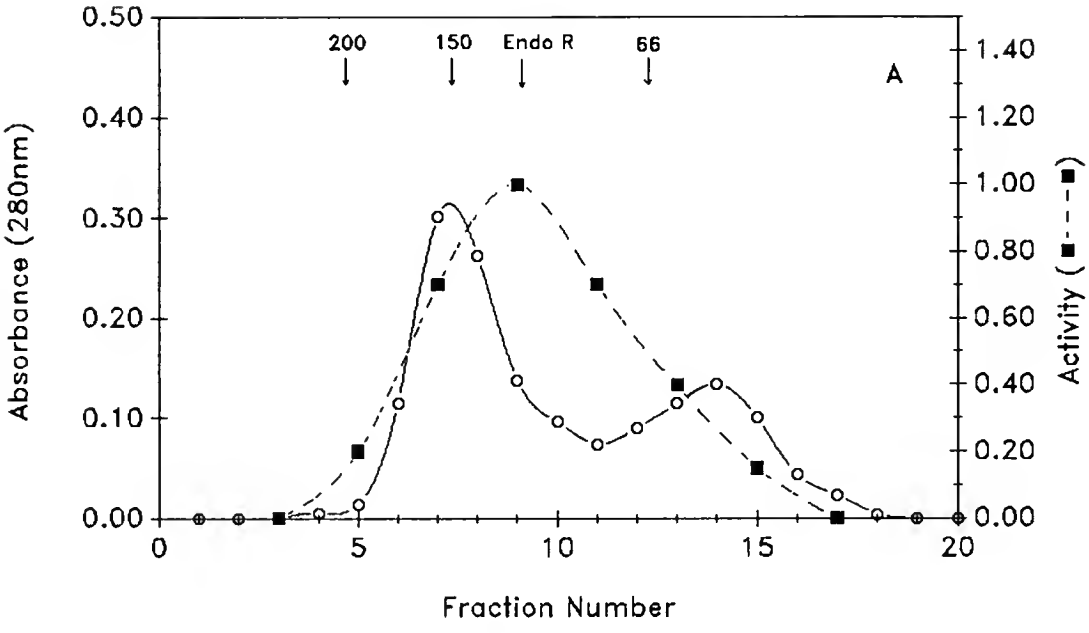

.13
.25
.50
1.0
2.0
3.0
-II
III
-7.4
-6.1
-3.0
-1.7

glycerol gradient (Figure 3-4A), in which the peak of double-stranded cleavage activity sedimented slightly more slowly than the 150 kdal marker. The sedimentation results correspond to an empirically determined sedimentation coefficient ($S_{20,W}$) of 6.8 (Figure 3-4B) and a molecular weight of 115 kd, assuming a spherical structure of the enzyme.

Pooled fractions from the poly(dG) agarose column, fraction V (Figure 3-5A, fractions 15-21), contain one major and 4 minor protein bands when analyzed by SDS polyacrylamide gel electrophoresis and silver staining. However, when individual fractions across the peak of endo R activity were analyzed on SDS gels (Figure 3-5B) and compared with the activity profile in the agarose gel assay (Figure 3-5A), the pattern of the major 100,000 molecular weight band was most consistent with the profile of specific cleavage activity. In addition, two other proteins of minor intensity, with molecular weights of 56 and 89 kd, appeared in a subset of the active fractions in a pattern not inconsistent with the activity profile (Figure 3-5C). As a result, it is possible that these peptides are responsible for the cleavage activity or contribute either as minor subunits or accessory peptides. It is also possible that the 56 and 89 kd peptides are proteolytic breakdown products of the major 100 kd protein. Attempts to purify endo R further or to isolate an active form from a non-denaturing acrylamide gel have been unsuccessful, probably due to the instability of the protein at low concentrations or to the loss of subunits and/or accessory proteins.

Figure 3-4. Sedimentation Analysis of Endo R by Glycerol Gradient Centrifugation.

A. Glycerol gradients (20-40%), in 0.1 M Tris-HCl, pH 7.5, 0.1 mM DTT were prepared in 5 ml polyallomer ultracentrifuge tubes. One milligram of fraction III endo R was combined with 0.25 mg of alcohol dehydrogenase, in 0.2 ml 0.1 M Tris-HCl, pH 7.5, 0.1 mM DTT. The mixture was layered on top of a glycerol gradient and the tube was centrifuged for 30 hours at 45,000 rpm and 4^{o}C in a Beckman SW 50.1 rotor. Three other gradients containing endo R fraction III alone, a mixture of the molecular weight standards B-amylase (200 kd) and BSA (66 kd), or alcohol dehydrogenase (150 kd) were run in parallel. Fractions were collected from the bottom of the tube and the absorbance at 280nm was determined (open circles). Fractions from gradients containing endo R were also assayed for double-stranded cleavage activity using the standard excision assay (filled squares). Sedimentation is from right to left. B. The sedimentation coefficient ($S_{20,W}$) of endo R was determined from a plot of the mobility of each protein, represented as a fraction of the total gradient length, against the $S_{20,W}$ literature values (CRC Handbook of Biochem., 1979) of the following protein standards: B-amylase (sweet potato, 8.9), alcohol dehydrogenase (yeast, 7.61), bovine serum albumin (4.41).


200
150
Endo R
66
A
0.50
0.40
0.30
0.20
0.10
0.00
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
0
5
10
15
20
Absorbance (280nm)
Activity (■---■)
Fraction Number

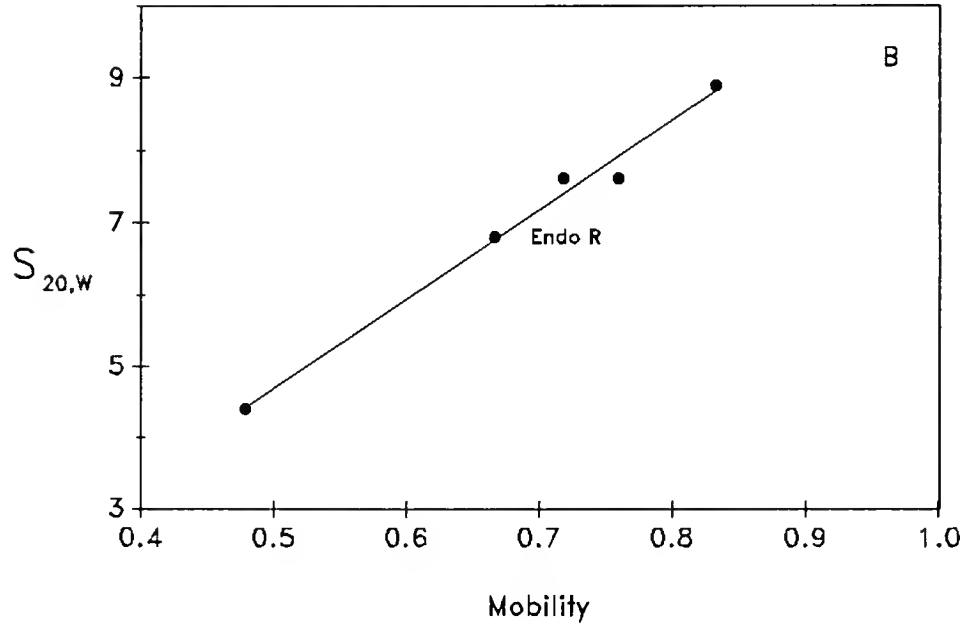
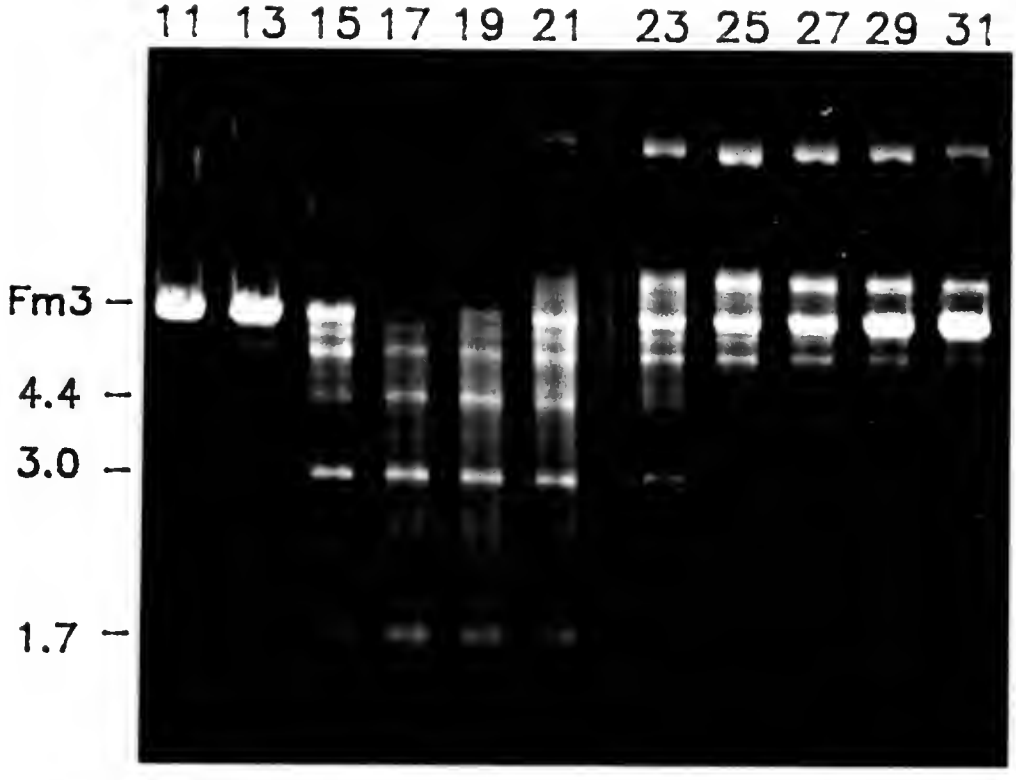

B
9
7
5
3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Endo R
S 20,W
Mobility

Figure 3-5. Molecular Weight and Purity of Endo R.

A) Agarose gel assay of peak fractions from poly(dG) agarose (Fraction V). Reaction mixtures of 25 μl containing 0.1 pmols of BstEII digested pSM620 DNA were incubated with 5 μl of the fraction indicated under standard reaction conditions. The reactions were stopped after 1 hour with the addition of 1/10 volume of gel running dye containing 0.2 M EDTA and 1% SDS and then fractionated on a 1.4% agarose gel. B) SDS polyacrylamide gel electrophoresis of peak fractions from poly(dG) agarose. A constant volume (20 μl) of the fractions indicated were added to an equal volume of 2X sample buffer, heated to 100°C for 3 minutes and separated on an SDS acrylamide gel as described in Chapter II: Materials and Methods. A parallel slot contained the following protein standards: B-amylase (205 kd), alcohol dehydrogenase (116 kd), phosphorylase b (97 kd), bovine serum albumin (66 kd), egg albumin (45 kd), and carbonic anhydrase (29 kd). C) Molecular weights of the peptides in Fraction V endo R. Poly(dG) agarose fractions 15-23 were pooled to form fraction V and the molecular weights of the protein species present were determined from their relative mobilities on an SDS gel. Open circles represent the molecular weight standards described for B; filled circles are endo R fraction V proteins.


11 13 15 17 19 21 23 25 27 29 31
Fm3
4.4
3.0
1.7

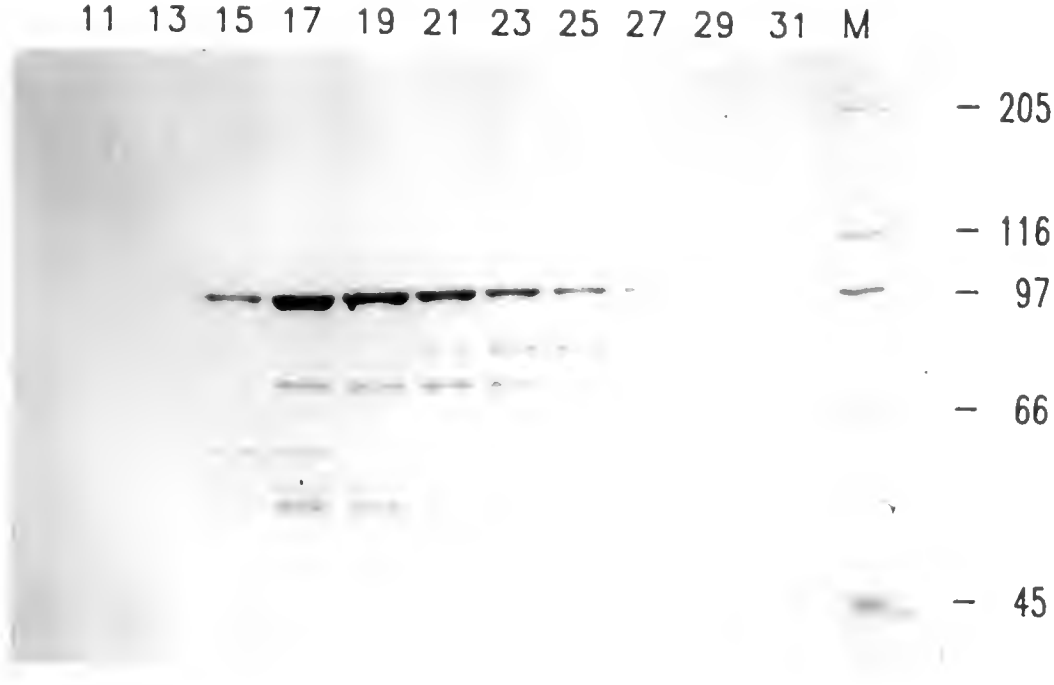


Figure 3-5 (continued), part B

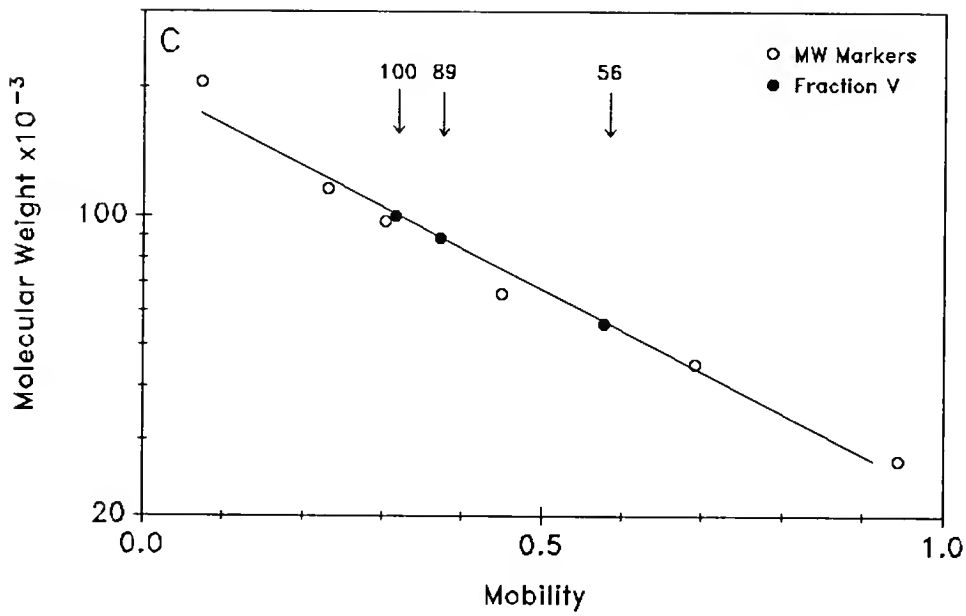


Figure 3-5 (continued), part C

CHAPTER IV
RESCUE OF AAV SEQUENCES

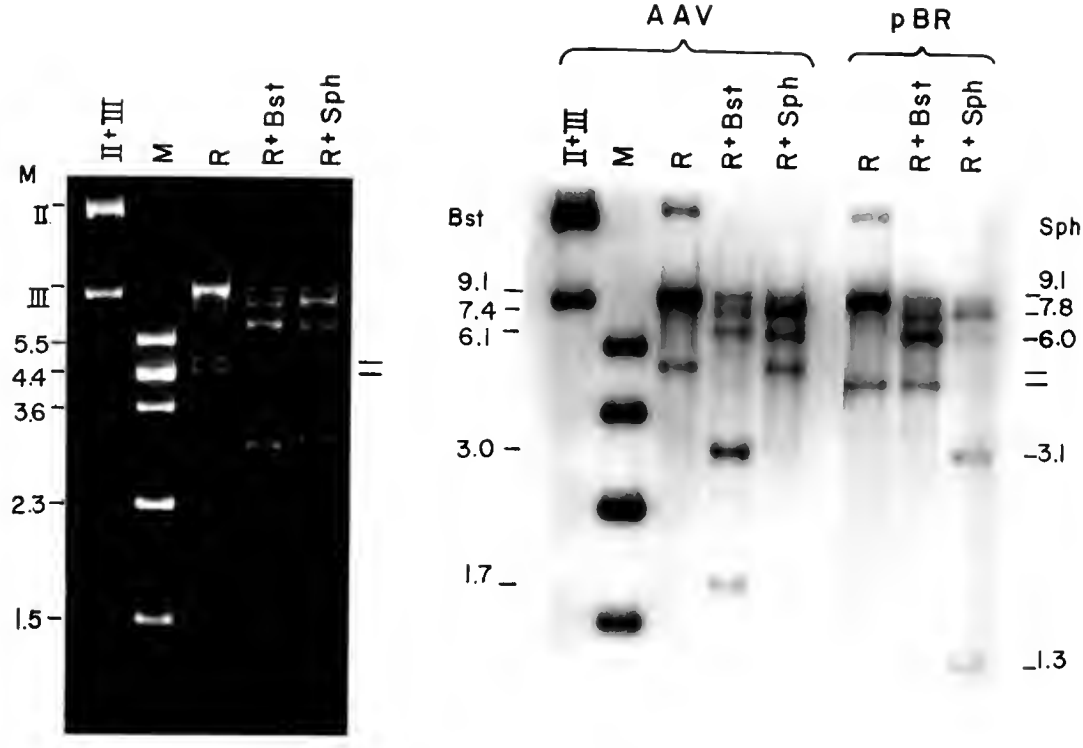
Localization of the Site of Cleavage to the AAV/Vector Junction

The excision of AAV sequences by a site specific endonuclease must be rather precise in order to produce substrate forms that are capable of replication without further modification. While minor deletions of the terminal sequences incurred during rescue can be repaired through gene conversion (Samulski et al., 1983), a major deviation in the site of cleavage would produce replicative forms that either lacked the ability to self-prime, or contained extraneous vector DNA that would have to be further processed prior to DNA replication. Therefore, an enzyme postulated to be responsible for AAV rescue in vivo must take these constraints on the site of cleavage into account.

To demonstrate that the site of cleavage was at the AAV/pBR322 junction, the products of the endo R reaction were hybridized to AAV- and pBR322-specific probes (Figure 4-1). Specific cleavage occurring precisely at the AAV/vector junction would produce a subset of fragments that consisted of AAV or pBR322 sequences exclusively, and these could be identified by probing duplicate Southern blots (Southern, 1975) with AAV or pBR322 DNA (Figure 4-1, right panel). In this experiment, reaction conditions were chosen to cleave 90% of the substrate (form I pSM620) at least once. When the products of the reaction were examined by ethidium bromide staining without further digestion with a restriction enzyme, most of the product formed was linear plasmid DNA

Figure 4-1. Identification of AAV-Specific and pBR-Specific Fragments in the Endo R Reaction.

PSM620 substrate was incubated with fraction IV enzyme in a standard endo R reaction (R). Where indicated the reaction products were also digested with BstEII or SphI. The products were fractionated in duplicate on 1.4% agarose gels, transferred to nitrocellulose filters, and hybridized to either AAV-specific or pBR-specific probe (right panel). The endo R plus BstEII and endo R plus SphI fragment sizes are indicated to the left and right of the right panel, respectively. The left panel illustrates the reaction products stained with ethidium bromide. Also indicated in the left panel are the sizes and positions of the marker bands (M) and pSM620 relaxed circular (II) and linear (III) plasmid DNA. The two unidentified lines to the right of each panel indicate the position of linear AAV (upper) and linear pBR (lower) DNA.


II+III
M
R
R+Bst
R+Sph
M
II
III
5.5
4.4
3.6
2.3
1.5
II
AAV
pBR
Bst
Sph
9.1
7.4
6.1
3.0
1.7
9.1
7.8
6.0
3.1
1.3

(Figure 4-1, left panel, lane R). In addition, approximately 3% of the products consisted of two fragments, 4.7 and 4.4 kb in length, which were the sizes expected for excised duplex AAV and linear pBR322 DNA. These were produced from double-stranded cleavage at both of the AAV/pBR junctions. The identity of these fragments was confirmed by Southern hybridization with selective probes. The slower migrating 4.7 kb fragment hybridized exclusively to AAV-specific probe and the 4.4 kb linear fragment hybridized only to pBR-specific probe (Figure 4-1, right panel, R lanes). The fragments generated by further digestion of the endo R products with either BstEII or SphI (see Figure 3-1) produced a set of fragments consistent with endo R cleavage at either one or the other of the AAV/pBR322 junctions. The 3.0 and 1.7 kb fragments produced from BstEII digestion of endo R products contained only AAV sequences and, therefore, hybridized only to AAV-specific probe. Conversely, the digestion of endo R products with SphI generated 3.1 and 1.3 kb fragments, which consist entirely of pBR322 DNA, and were detected only with pBR322-specific DNA probe. As expected, the higher molecular weight fragments contain both AAV and pBR sequences and, therefore, hybridize to both probes. These results, limited by the sensitivity of hybridization, placed the position of the cleavage site to within 25 bp of either AAV/vector junction.

Either AAV/Vector Junction Can Be Cleaved Independently of the Other

One proposed mechanism for gene conversion of AAV terminal mutants to the wild type form requires the interaction between the two terminal repeats (Lusby et al., 1980; Samulski et al., 1983). It was therefore possible that cleavage by endo R required this type of interaction. To

determine if this was the case, subclones of pSM620, pGM620C and pGM620D were constructed which contained either the left or right terminal repeat of AAV, respectively (Figure 4-2C). In addition to the 145 bp terminal sequence, both plasmids contained approximately 350 bp of flanking AAV DNA. Both pGM620C and pGM620D were efficiently cleaved by endo R and produced fragments of the expected size and sequence composition when the endo R products were digested with SphI (Figure 4-2, A and B). It was, therefore, concluded that either AAV/vector junction could be cleaved independently of the other and that the recognition sequence was contained within the terminal 500 bp of AAV. Figure 4-2 also demonstrates that vector DNA, pBR322, is a poor substrate for endo R.

Cleavage of AAV Terminal Mutants

A number of mutant AAV clones with deletions in the terminal sequences were fortuitously produced during the course of constructing a viable AAV recombinant clone (Samulski et al., 1983). To identify the endo R recognition signal more precisely, the sequences at the AAV/vector junctions for several of the mutant AAV plasmids were determined and the relative frequencies of cleavage were compared (Figure 4-3 and Table II). The conclusion from these studies was that there were three sets of sequences near the AAV/vector junctions that could serve as endo R recognition sites. The first was a stretch of poly(dG) residues that was present at many of the junctions and was a result of plasmid construction by GC tailing. A second sequence, which was called the AAV recognition signal, consists of nucleotides 3 to 15 of the AAV terminal repeat and contains the sequence CCaCTCCCTCTCT.

TABLE II

RECOGNITION SITES FOR ENDO R
IN AAV PLASMIDS[a]

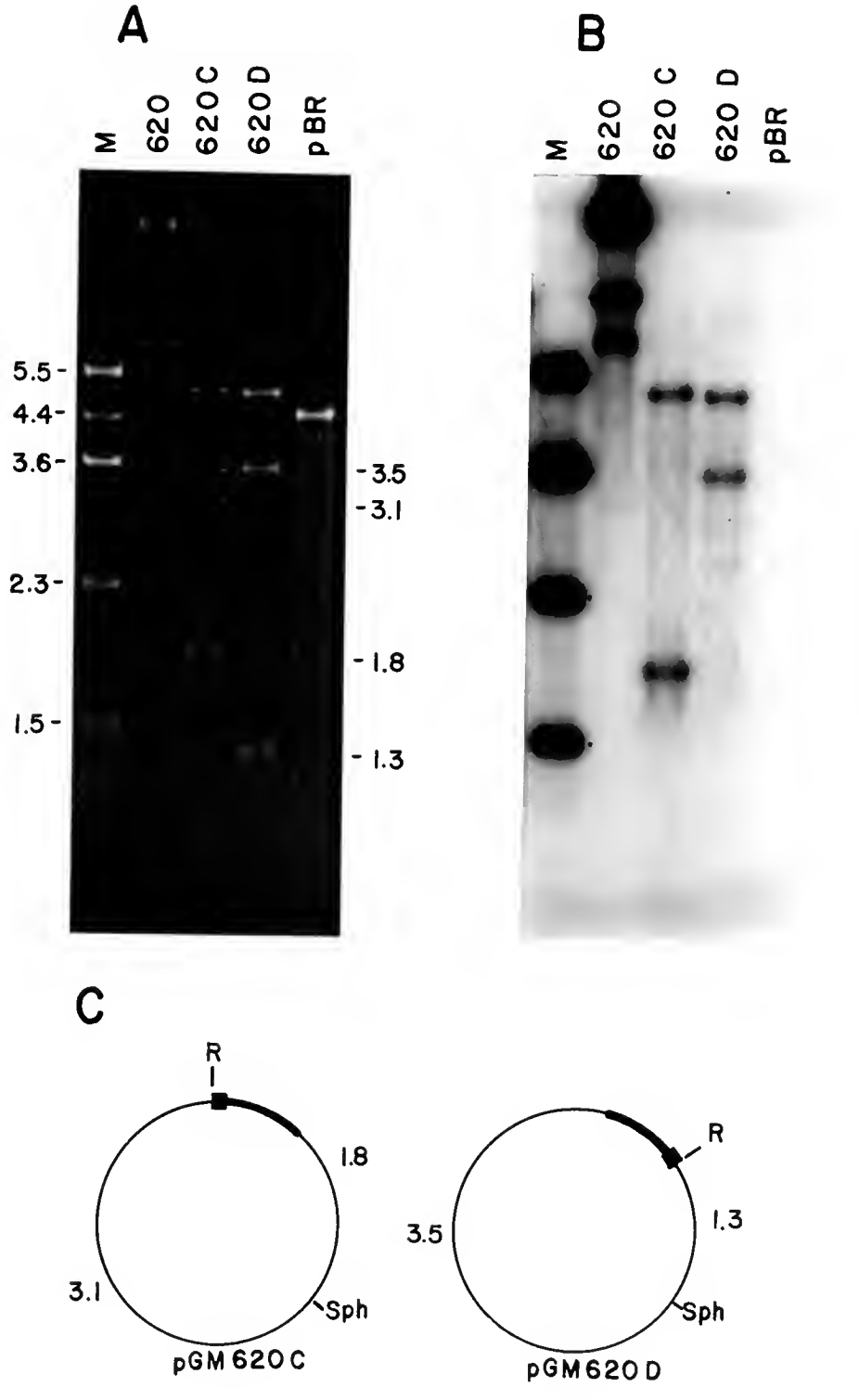
Junction				Relative Frequency[b]
(g)n + CCaCTCCCTCTCT				
pSM620(L)	Vector-	(g)19	CCaCTCCCTCTCTgCgCg -AAV	100
pSM620(R)	Vector-	(g)29	CCaCTCCCTCTCTgCgCg -AAV	100
pSM621(R)	Vector-	(g)29	CCaCTCCCTCTCTgCgCg -AAV	100
pSM703(R)	Vector-	(g)38	CCaCTCCCTC|aCTaggg -AAV	150
pSM609(R)	Vector-	(g)10	|CTCCCTCTCTgCgCgCTC -AAV	50
CCaCTCCCTCTCT				
pSM621(L)	AAV-		CCaCTCCCTCTCTgC|agC -Vector	1
pSM704(L)	AAV-		CCaCTCCCTCTCTgC|agC -Vector	1
pSM609(L)	AAV-		CCaCTCCCTCTCTgC|agC -Vector	1
(g)n				
pSM703(L)	Vector-	(g)36	|agTggCCaaCTCCaTCaC -AAV	150

[a]The table lists the sequences at the AAV/vector junctions and the frequency of cleavage of the AAV clones used to determine the recognition site for endo R. The first two bases of the AAV sequence in pSM620 (gg) are counted as part of the poly(dG):poly(dC) tail. A vertical line (|) indicates the position of the deletion in a mutant clone. pSM704(R) and both pAV1 plasmid junctions were not sequenced. (R) and (L) depict the right and left AAV/vector junctions, respectively. pSM620 and pSM609 had been previously sequenced (Samulski, et al., 1983) and the sequence of pSM609(L), which was found to be incorrect, is changed here. The remaining plasmids were sequenced as part of this study.

[b]Indicates the approximate yield of the fragment resulting from cleavage at the indicated site (Figure 4-3), where 100% cleavage has been defined as the cleavage seen at pSM620(R) No attempt was made to distinguish between the level of cleavage observed at the junctions in each category.

Figure 4-2. Cleavage of Substrates That Contain Only One Copy of the AAV Terminal Repeat.

C) Restriction maps of pGM620C and pGM620D. pGM620C contains the 513 bp left terminal PstI fragment of the wild type AAV plasmid pSM620, including the left AAV terminal repeat and the 19 bp poly(dG) tail. PGM620D includes the right terminal repeat of AAV and the 29 bp GC tail in a 445 bp PstI fragment subcloned from pSM620. Thick lines represent AAV DNA, thin lines are pBR DNA, and the solid boxes are the AAV terminal repeats. Also indicated are the positions of the endo R (R) and SphI sites. A and B) 0.5 μg of pGM620C, pGM620D or pBR322 supercoiled plasmid DNA were incubated with fraction III enzyme as described in the legend for Figure 3-3 and digested with SphI. The reaction products were fractionated in a 1.4% agarose gel, stained with ethidium bromide (panel A), transferred to a nitrocellulose filter and hybridized to nick-translated AAV DNA probe (panel B). The position of the marker bands (M) and the SphI digested endo R product bands are indicated to the left and right of panel A, respectively. The 620 lanes contain forms I, II, and III pSM620 plasmid DNA.

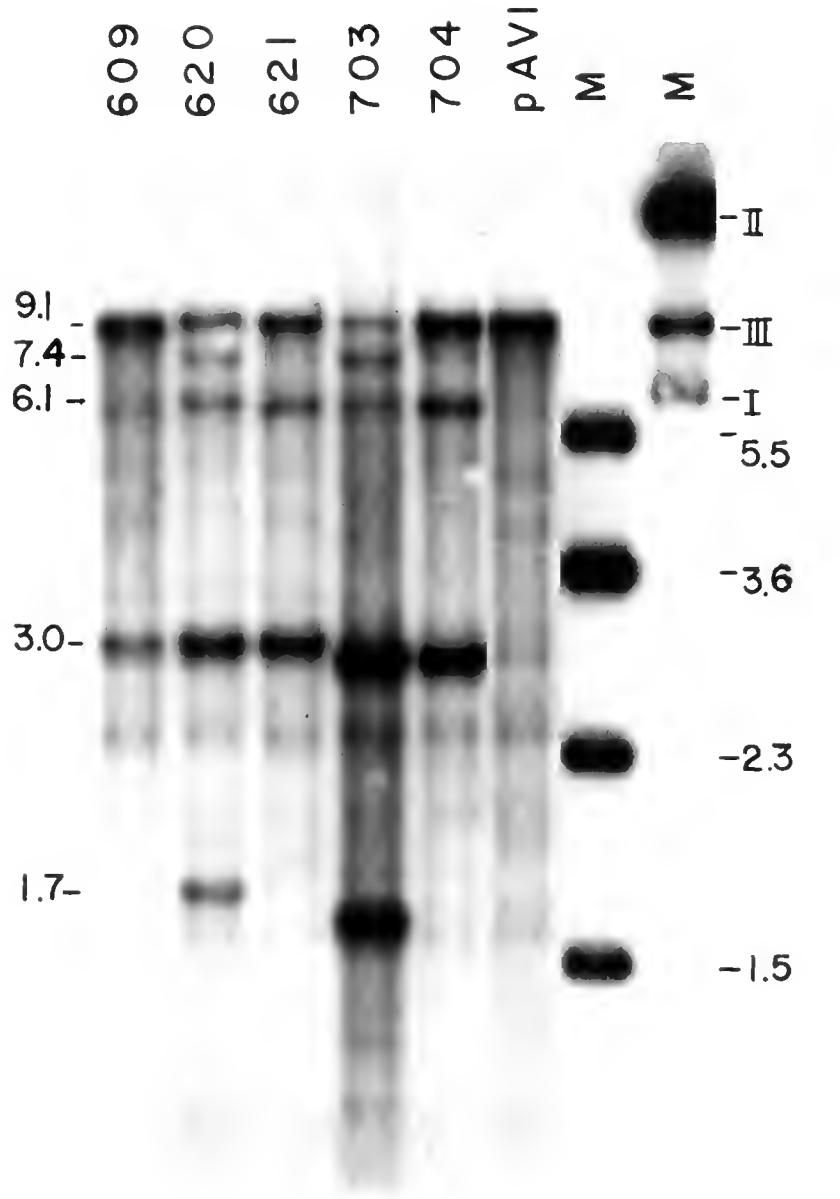

A
M
620
620C
620D
pBR
5.5-
4.4-
3.6-
2.3-
1.5-
-3.5
-3.1
-1.8
-1.3
B
M
620
620 C
620 D
pBR
C
R
1.8
3.1
Sph
pGM620C
R
3.5
1.3
Sph
pGM620D

Because the AAV terminal repeat is a palindrome, this sequence occurs twice within each wild type terminal repeat, at nucleotides 3 to 15 and 109 to 121 (Lusby et al., 1980). The two copies have been designated the outboard and inboard recognition sequences, respectively. The presence of the inboard copy of this sequence accounts for the minor cleavage band derived from the left end of AAV and seen at the 1.6 kb position in virtually all of the plasmids shown in Figure 4-3. It is essentially the 1.7 kb Endo R/BstEII fragment minus the first 100 bp of AAV DNA, which is the distance between the outboard and inboard copies of the AAV recognition sequence. In 3 junctions (pSM621(L), PSM704(L), and pSM609(L); Figure 4-3 and Table II), the AAV cleavage site was the only recognition sequence present because the deletion in these plasmids has removed the poly(dG) tail and placed the inboard copy of the sequence adjacent to vector DNA. Accordingly, the 1.6 kb band was the only fragment seen in these plasmids as the result of cleavage at the left end. Endo R cleavage of the left junction of the wild type clone, pSM620(L), produced both the 1.6 and 1.7 kb bands because both the inboard and outboard copies of the AAV recognition sequence were present. Presumably, the inboard site at the right end of AAV was also cleaved, but the separation in this region of the gel was not sufficient for resolution of the 2.9 kb fragment (Figure 4-3).

The relative intensities of the 1.7 and 1.6 kb band produced from pSM620(L) cleavage reflect the presence of the poly$(dG)_{19}$ sequences adjacent to the outboard site. In general, the homopolymeric poly(dG) tail sequence was approximately 20 to 100 fold more likely to be cut than the heteropolymeric AAV recognition sequence (Table II). In contrast to pSM620, another wild type AAV clone, pAV1, contains no GC

Figure 4-3. Cleavage of AAV Plasmids That Contain Deletions in the Terminal Repeats.

The various AAV mutant substrates were incubated with fraction III enzyme and digested with BstEII as described for Figure 3-3. The reaction products were transferred to nitrocellulose and probed with ^{32}P-labeled AAV DNA. The molecular weights of the endo R/BstEII products are indicated at the left of the figure. The sequence at each AAV/pBR322 junction is listed in Table II. The figure was over-exposed to visualize minor cleavage bands.


609
620
621
703
704
pAV1
M
M
9.1
7.4
6.1
3.0
1.7
II
III
I
5.5
3.6
2.3
1.5

tail and cleavage was directed exclusively by the AAV terminal recognition site. As a result, cleavage at the inboard and outboard sites occurs with approximately equal efficiency to produce two junction bands from each end of equal intensity (Figure 4-3, pAV1; the fragment sizes are 3.0, 2.9, 1.7 and 1.6 kb). The fact that poly(dG)$_n$ alone was sufficient to act as a recognition sequence was further demonstrated by cleavage of pSM703(L). This plasmid contains a deletion which has removed both the inboard and outboard AAV recognition sites, but retains a poly(dG) stretch that is 36 bp long. The poly(dG) stretch was sufficient to promote cleavage and produced a band of approximately 1.6 kb. The slightly slower mobility of the 1.6 kb fragment generated from pSM703(L) cleavage is attributed to the addition of the sizable poly(dG) tail.

Minor Endo R Cleavage Sites

Cleavage by endo R appeared to require a sequence that was at least 10 bp long and consisted of a polypurine-polypyrimidine tract that was relatively rich in GC base pairs. These requirements were confirmed by the identification of some of the minor cleavage fragments observed in Figure 4-3, which had been predicted on the basis of this sequence. It was noticed that a number of minor bands appeared consistently in digests of all of the AAV variant plasmids. Because the size of the minor fragments did not vary with the size of the terminal deletion, these fragments must have been generated from cleavage within AAV sequences (If the fragments had been generated from cleavage within pBR322, they would have spanned the pBR/AAV junction and their size would have varied depending on the size of the terminal deletion).

TABLE III

MINOR CLEAVAGE SITES[a]

	Computer-Predicted Fragments			Observed
Sequence	Nucleotide	kb	%C	kb
CCCCTCTCCCCTC	4089	2.39	10/13	2.4
CCaCTCCCTCTCT	4678	3.00	8/13	3.0
CCaCTCCCTCTCT	4562	2.86	8/13	NC
CCaCTCCCTCTCT	3	1.70	8/13	1.7
CCaCTCCCTCTCT	121	1.59	8/13	1.6
CTCCaCCCCTCC	157	1.54	9/12	NC
CTaCagCaCCCCTT	3026	1.33	7/14	1.3
CCCTgCCCaCCT	2947	1.25	8/12	1.2
CagCagCCCCCTCT	2775	1.08	8/14	1.1
CCagaCTCCTCCTC	2661	0.96	8/14	ND
CCCgCCTCCggCgCC	753	0.95	10/15	ND
CCCCTCCTCCCaCC	1461	0.24	11/14	TS

[a]The table lists the cleavage fragments predicted by a computer search of the plasmid pSM620 using poly(dC)$_{13}$ as the search sequence and allowing for 60% homology between compared sequences. The sequence at which cleavage would presumably occur and the number of C residues in each sequence (%C) are tabulated with the size of the predicted fragment, which was derived by assuming that endo R cleavage occurred 3′ to the poly(dG) sequence. The predicted fragment sizes are compared with the sizes of the fragments observed in the agarose gel assay. All fragment sizes are the result of BstEII digestion of endo R products. NC = Not clear; ND = Not detected; TS = Too small to be resolved on a 1.4% agarose gel.

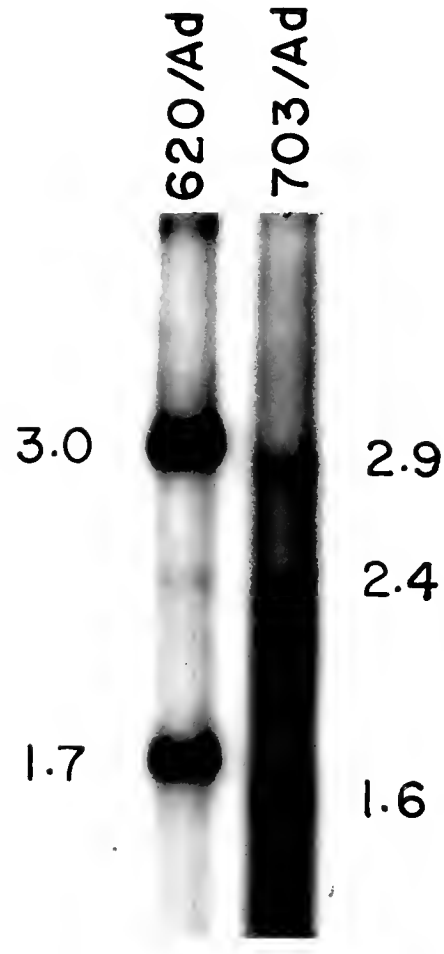
Using a computer and allowing for substitutions within the predicted consensus sequence, poly(dG)$_{19}$, or the complementary sequence poly(dC)$_{19}$, internal AAV sequences were searched for potential endo R recognition sites. Table III compares the computer predicted cleavage sites and fragments with the observed fragments. The results show a good correspondence between the computer-predicted cleavage pattern and the fragments seen in endo R digests. In fact, all of the observed fragments that were believed to be the result of cleavage in AAV DNA were predicted by the computer search. In particular, the best matches found were with a sequence at nucleotide 4089, which produced the 2.4 kb minor cleavage fragment, and with both the inboard and outboard AAV terminal recognition sequences.

AAV Excision In Vivo

A demonstration that endo R was capable of excising AAV sequences *in vivo* in a manner similar to that observed *in vitro* would add credibility to the hypothesis that this was the enzyme involved in AAV DNA replication. When AAV plasmids were transfected into tissue culture cells, the bulk of the input DNA was found in the form of nicked circular (form II) or linear (form III) plasmid species (Samulski *et al.*, 1982, 1983; Hermonat and Muzyczka, 1984). This was true regardless of whether the input AAV DNA was capable of DNA replication. To determine if the linear plasmid DNA (form III) produced *in vivo* was the result of cleavage at a specific sequence, transfected DNA was isolated after 21 hours, digested with BstEII and hybridized with AAV probe (Figure 4-4). The 1.7 and 3.0 kb bands were readily seen in cells infected with adenovirus and transfected with the wild type plasmid,

Figure 4-4. Cleavage of AAV Plasmids *In Vivo*.

Five micrograms of PSM620 and pSM703 DNA were transfected into Ad-infected HeLa cells (Chapter II: Material and Methods). At 21 hours after transfection, low molecular weight DNA was isolated by the method of Hirt (1967), digested with BstEII, fractionated by agarose gel electrophoresis and hybridized to AAV-specific probe as described for Figure 3-3.


620/Ad
703/Ad
3.0
2.9
2.4
1.7
1.6

pSM620. In this case the fragments were presumably derived from replicating AAV DNA. However, bands of the same size were seen at reduced levels in cells transfected with pSM620 in the absence of adenovirus or in the presence of adenovirus and hydroxyurea (data not shown). In addition, two minor bands were seen in cells transfected with pSM620 which apparently were not amplified by DNA replication. These were identical to the minor 2.4 kb and 1.6 kb bands observed *in vitro*.

Cleavage fragments generated from pSM703 transfection were expected to be approximately 100 bp shorter than those seen with the wild type plasmid due to the deletions in both termini (compare with Figure 4-3). These bands were readily seen in cells transfected with pSM703 DNA (Figure 4-4) and because pSM703 is an ori$^-$ mutant (Samulski *et al.*, 1983), the 1.6 and 2.9 kb fragments must have been produced from the input plasmid sequences. As expected, pSM703 also generated a minor 2.4 kb band. It was concluded, therefore, that a substantial amount of input plasmid DNA was cleaved at the AAV/vector junction *in vivo* and that AAV DNA replication is not necessary for cleavage to occur. Moreover, the pattern of cleavage was essentially the same *in vivo* and *in vitro*. The remaining input DNA was apparently cleaved randomly as judged by the background radioactivity.

Replication of Endo R Products

When the products of an *in vitro* endo R digestion of pSM620 were transfected into Ad-infected HeLa cells, DNA replication of the AAV sequences occurred at levels that were indistinguishable from the transfection of the form I plasmid (not shown). This was true even at

times early after transfection. Differences in the production of replicated AAV DNA would presumably occur early in the replication cycle, but only if the excision of AAV sequences was a rate limiting step in DNA replication. Therefore, endo R products were substrates for DNA replication *in vivo*, but cleavage of AAV plasmids *in vitro* prior to transfection did not provide a selective advantage for DNA replication.

Endo R products of pSM620 cleavage were also substrates for DNA Polymerase I *in vitro*. The products of endo R digestion were heat denatured and cooled rapidly to favor the formation of the hairpin primer over renaturation of duplex molecules. In these experiments, a significant amount of radioactivity was incorporated when these substrates were incubated with DNA Pol I (not shown). Thus, the products of endo R cleavage will support DNA replication both *in vitro* and *in vivo*.

CHAPTER V
ENZYME RECOGNITION

Isolation of the Cleavage Site

In previous experiments, a consensus cleavage site was defined on the basis of commonly occurring sequences found in AAV recombinant plasmids. These sequences contained homopurine-homopyrimidine stretches of at least 10 nucleotides which were rich in G:C base pairs. However, it remained possible that the endo R recognition signal was at a separate location from the cleavage site, possibly at an internal AAV sequence. In this case, enzyme recognition would occur at a distant site and result in cleavage at the nearest available stretch of homopurine-homopyrimidine sequences. To rule out this possibility and to further characterize the recognition sequence, several clones containing subsets of the sequences present at the AAV/vector junctions were constructed and their ability to act as substrates for endo R specific cleavage was evaluated. The level of cleavage in several of these clones is illustrated in Figures 5-1 and 5-2 and the insert sequence and frequency of cleavage are listed in Table IV. The first plasmid, pGM1008, is a subclone of pGM620D and contains sequences from the right AAV/pBR322 junction, including the first 21 bp of AAV and the 31 bp G:C tail. This plasmid was cleaved with approximately the same frequency as the parental junction in pSM620D (Figure 4-4). Two other plasmids, pGM913 and pGM1116, contained poly(dG)-poly(dC) stretches of 9 and 13 bp, respectively, inserted at the PstI site of pBR322.

Figure 5-1. Oligomer Clone Excision

A) Standard reaction mixtures of 25 μl, containing 0.4 pmol of the pGM clones indicated, were digested with endo R at 37°C for 1 hour. The reactions were terminated with the addition of proteinase K stop solution (see Chapter II: Materials and Methods), phenol extracted, ethanol precipitated and digested with EcoRI. B) EcoRI digestion of endo R cleaved substrate produces 0.8 and 3.6 kb bands from pGM clones containing oligomer inserts in the PstI site of pBR322, while 1.0 and 3.4 kb fragments are produced from clones which contain inserts in the EagI site (see Figure 5-6 for the sequence of the inserts). E = EcoRI site, R = endo R site, filled square = oligonucleotide insert (See Chapter II for a description of the clones).

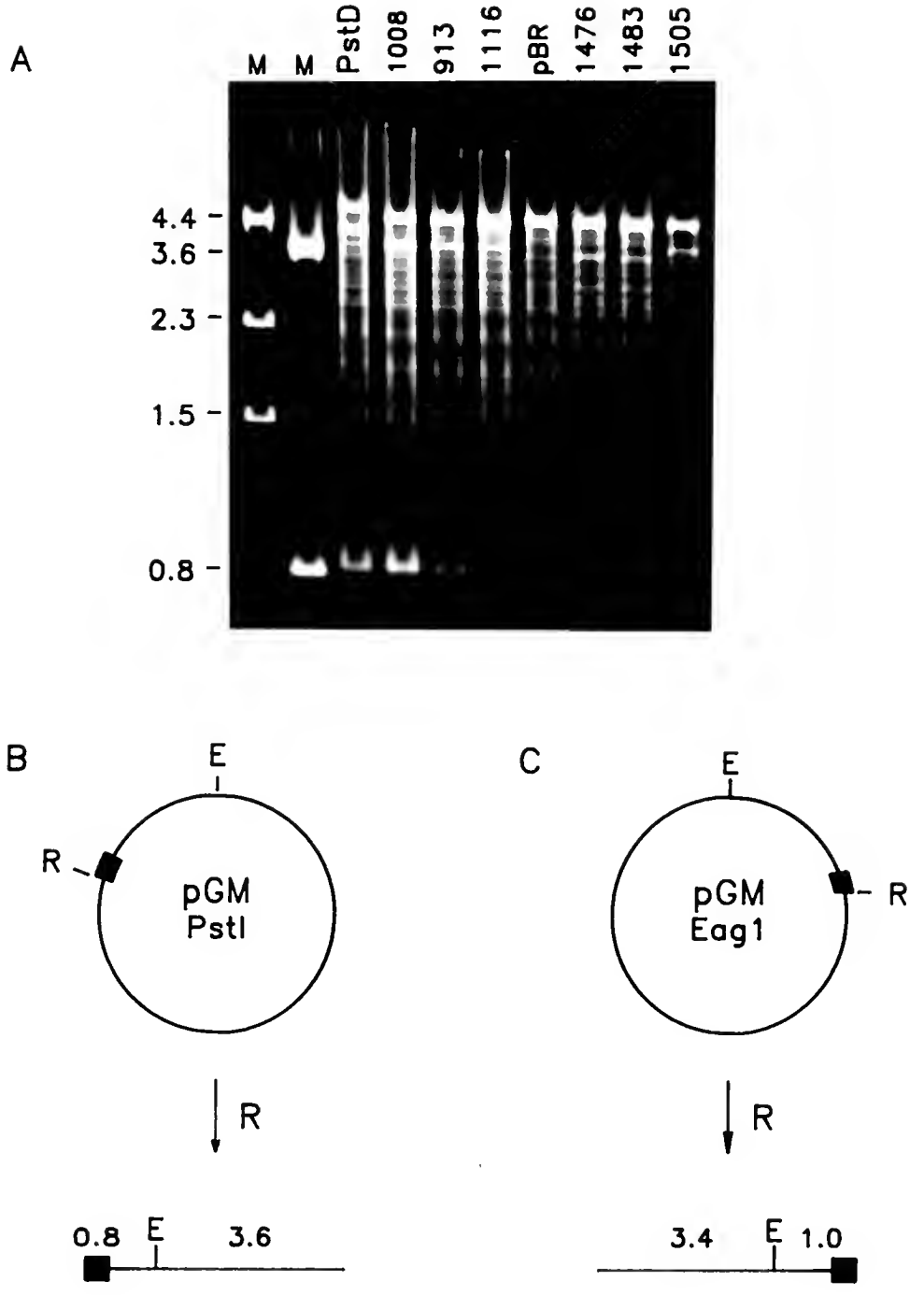

A
M
M
PstD
1008
913
1116
pBR
1476
1483
1505
4.4
3.6
2.3
1.5
0.8
B
E
R
pGM
PstI
R
0.8
E
3.6
C
E
pGM
Eag1
R
R
3.4
E
1.0

Figure 5-2. Oligomer Clone Gel Assay.

Standard reactions of 25 μl, containing 0.4 pmol of the substrates indicated, were digested with endo R at 37^{o}C for 1 hour. Reactions were terminated with proteinase K stop solution (Chapter II), phenol extracted, ethanol precipitated and digested with EcoRI. The assays were fractionated on a 1.7% agarose gel, transferred to nitrocellulose (Southern, 1975) and probed with nick translated pBR322 DNA (>1 x 10^{8} cpm/μg). Light and dark exposures (top and bottom, respectively) of the 0.8 kb fragments produced from endo R and EcoRI digestion are shown to illustrate the relative frequency of cutting in the different substrates.

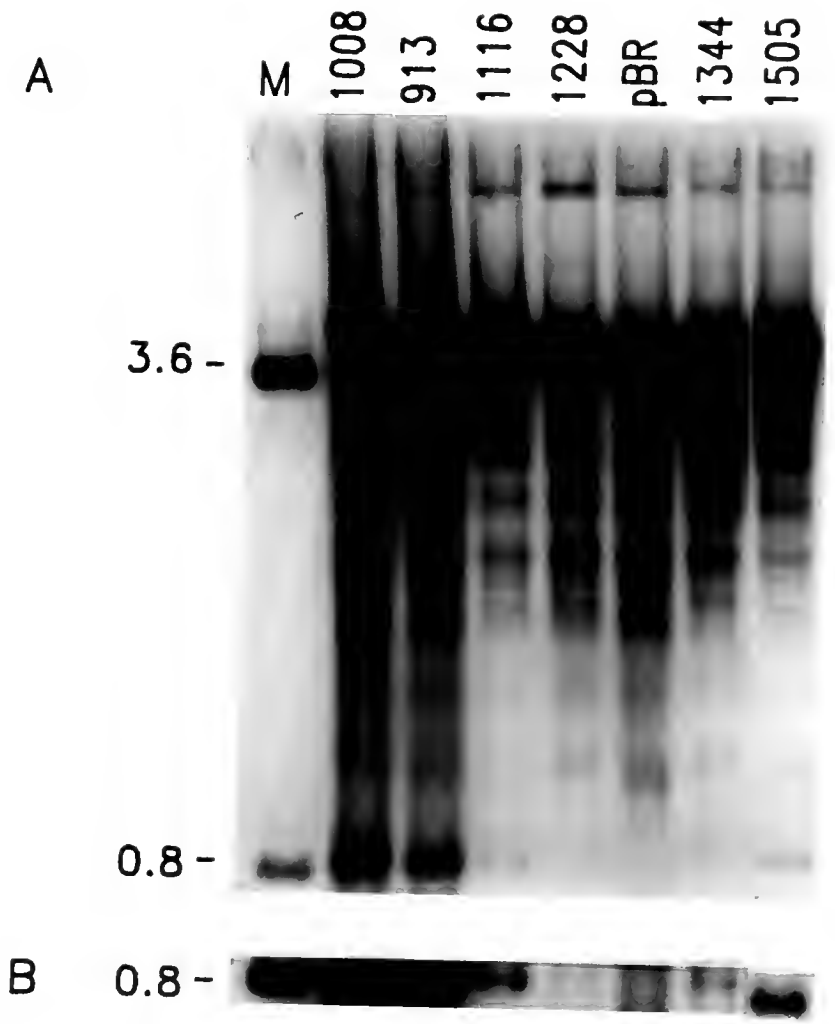

A
M
1008
913
1116
1228
pBR
1344
1505
3.6 -
0.8 -
B
0.8 -

TABLE IV

SUBCLONED CLEAVAGE SITES[a]

Clone	Inserted Sequence[b]		Frequency
pGM1008	$(g)_{29}$	CCaCTCCCTCTCTgCgCg	100
pGM913	$(g)_{13}$		30
pGM1116	$(g)_{9}$		20
pBR322[c]	$(g)_{6}$		0
pGM1228	$(g)_{2}$	CCaCTCCCTCTCTgCgCg	1
pGM1344[d]	$(g)_{2}$	$(CCaCTCCCTCTCTgCgCg)_{2}$	5
pGA38	$(ga)_{38}$		10
pGA11	$(ga)_{11}$		3
pGM1635	$(gC)_{20}$		0
pGM1483	$(C_4a_2)_3$		30
pGM1505	$(C_{2-6}T)_5$		30

[a]The construction of the clones is described in detail in Chapter II: Materials and Methods. The sequences are listed as they appear in the clone on the clockwise strand in a 5′ to 3′ direction. The frequency of cleavage is defined as the percent of starting substrate cleaved in a standard assay with 1 unit of endo R. The amount of cleavage for pGM1008 was arbitrarily set to 100%.

[b]The complete sequence of the inserts are listed in Figure 5-6.

[c]Natural poly(dG) stretches of 6 bp occur in pBR322 at nucleotides 2550 and 2797.

[d]The sequences for pGM1344 are repeated in an inverted orientation

No AAV sequences were present in either plasmid. Specific cleavage was observed with these substrates as well. However, the level of cleavage was significantly reduced with both substrates. PGM913 (13 bp of G:C) was cleaved at 30% and pGM1116 (9 bp of G:C) was cleaved at 20% of the level observed with clones that contained all of the parental AAV/vector junction (pGM620D, pGM1008, Figures 5-1 and 5-2, Table IV). These clones approach the limit of enzyme recognition, since the naturally occurring stretches of poly(dG)$_6$ in pBR322 (Figures 5-1, 5-2 and Table IV) and poly(dG)$_7$ in lambda phage DNA (not shown) were not detectably cleaved. These results, in conjunction with those obtained with the pSM609(R) junction (Table II), suggested that a minimum of 9 G residues are sufficient for cleavage, while the frequency of cleavage increases in direct proportion to the length of the homopolymer chain. Thus, AAV plasmids that contain GC tails at the AAV/vector junction are ideal substrates for endo R.

The apparent strong affinity of endo R for long stretches of poly(dG) and the observed cleavage of G-rich sites that did not contain homopolymers of G:C (Figure 4-3, pAV1, pSM609(L), pSM621(L) and pSM704(L)) raised some important questions concerning alternative enzyme recognition sites and, particularly, what effect substitutions in the poly(dG) chain have on cleavage activity. To determine what type of substitutions were tolerated by endo R, a number of plasmid substrates containing a variety of repeating polymeric sequences were tested. The first of these was the plasmid pEV136, an infectious polio clone which contains 18 bp of poly(dG) at the 5′ end and an 84 bp stretch of poly(dA) at the distal end. When incubated with endo R, pEV136 produced fragments which were the result of cleavage exclusively at the poly(dG)

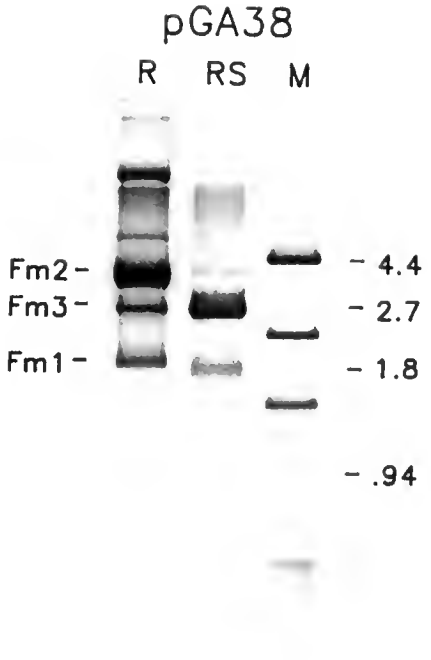
site. No detectable cleavage was observed at the poly(dA) site (Figure 5-3). However, clones that contained an insert of the alternating copolymer $(GA)_{38}$ and $(GA)_{11}$ (pGA38 and pGA11, respectively) were cleaved by endo R at a reduced frequency (Table IV and Figure 5-3). Thus, homopolymers of (dA) were not recognized as substrates for cleavage, but substitutions of (dA) in long stretches of poly(dG) were tolerated and resulted in a decrease in activity. This was not the case with alternating polymers of GC that can potentially form altered secondary structures of left-handed Z-DNA (Peck and Wang, 1983; Singleton et al., 1983). Clones containing inserts of $(GC)_{10}$ and $(GC)_{20}$ were not detectably digested by endo R (pGM1635, Table IV).

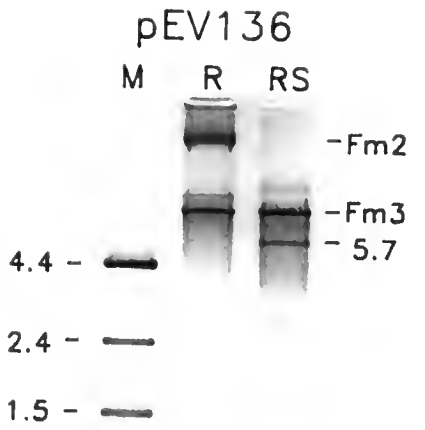
More complex sequences with less symmetrical substitutions of the poly(dG) sequence are also substrates for endo R. This is exemplified by two clones, pGM1483 and pGM1505, that contain generic forms of the Tetrahymena (C_4A_2) and Dictyostelium ($C_{2-6}T$) telomeric sequences, respectively (Blackburn and Szostak, 1984; Shampay et al., 1984). These substrates were cleaved at a level approaching the frequency observed with the homopolymeric clone pGM913 (Figure 5-1 and Table IV).

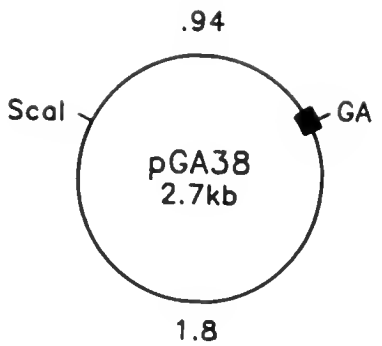
Integrated AAV provirus exists in two basic forms in chromatin, as well as in plasmids. The first of these consists of at least two tandem head to tail copies of AAV, each separated by two copies of terminal sequences in an inverted orientation (McLaughlin et al., 1988). Presumably, this would place two copies of the endo R recognition site in an inverted orientation, adjacent to each other. The plasmid equivalent of this contained one copy of the AAV terminal repeat adjacent to a GC tail and was cleaved with high frequency (e.g. pGM1008, Figure 5-1). The second type of integrated form occurs less frequently

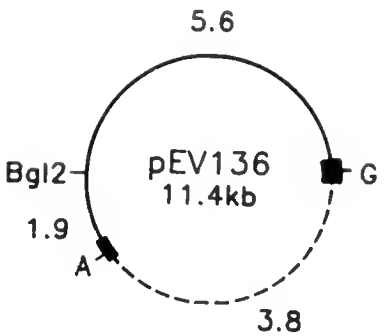
Figure 5-3. Endo R Cleavage of pGA38 and pEV136.

A) One microgram of either pGA38 or pEV136 was incubated with 1 unit of fraction V endo R under standard reaction conditions in a volume of 25 μl. After phenol extraction and ethanol precipitation, one half of each reaction was further digested with either ScaI (pGA38) or Bgl2 (pEV136) and both the endo R treated (R) and the restriction enzyme digested-endo R treated (RS) samples were fractionated on 1.4% agarose gels. B) Restriction maps of pGA38 and pEV136. Filled boxes represent the poly(GA) insert for pGA38 (GA) and the poly(dG) and poly(dA) inserts for pEV136 (G and A, respectively). The molecular weights (kb) indicate the size of the products that would be expected from a restriction enzyme digest of endo R products cleaved at GA in pGA38, and at G plus A in pEV136.


pGA38
R
RS
M
Fm2
Fm3
Fm1
4.4
2.7
1.8
.94


pEV136
M
R
RS
Fm2
Fm3
5.7
4.4
2.4
1.5


.94
ScaI
GA
pGA38
2.7kb
1.8


5.6
pEV136
11.4kb
Bgl2
G
1.9
A
3.8

in chromatin and contains single or multiple copies of AAV separated by a single copy of the AAV terminal repeat (McLaughlin et al., 1988). Plasmids that contain this type of arrangement were cleaved at a low frequency in vitro (e.g. pAV1, Figure 4-3).

To determine if the terminal 23 nucleotides of AAV contained a recognition signal for endo R and what effect the number of copies and orientation of these sequences had on activity, two clones containing either a single copy or two inverted copies of the AAV terminal recognition sequence (ggCCaCTCCCTCTCTgCgCgC) were constructed. The plasmid pGM1228 contains a single copy of this sequence and pGM1344 contains two inverted copies of the same sequence, both inserted into the PstI site of pBR322. Both of these constructs were cleaved by endo R (Figure 5-2). However, the level of cleavage observed with pGM1228 was approximately 1% of that observed with pGM1008 and could only be visualized by Southern hybridization. In contrast, the plasmid that contained two inverted copies of the recognition sequence was cut at a level at least 5 times greater than that of the single copy plasmid, while still only about 5% of that seen with the parental junction plasmids. Thus, the presence of tandem inverted copies of the AAV recognition signal improves the yield of specifically cleaved product in vitro. This arrangement may be representative of the preferred situation in chromatin.

Sequence at the Site of Cleavage

Previous experiments have already shown that endo R makes double-stranded specific cuts and that cleavage of the substrate occurs at the recognition signal. In addition, it was possible that cleavage at the

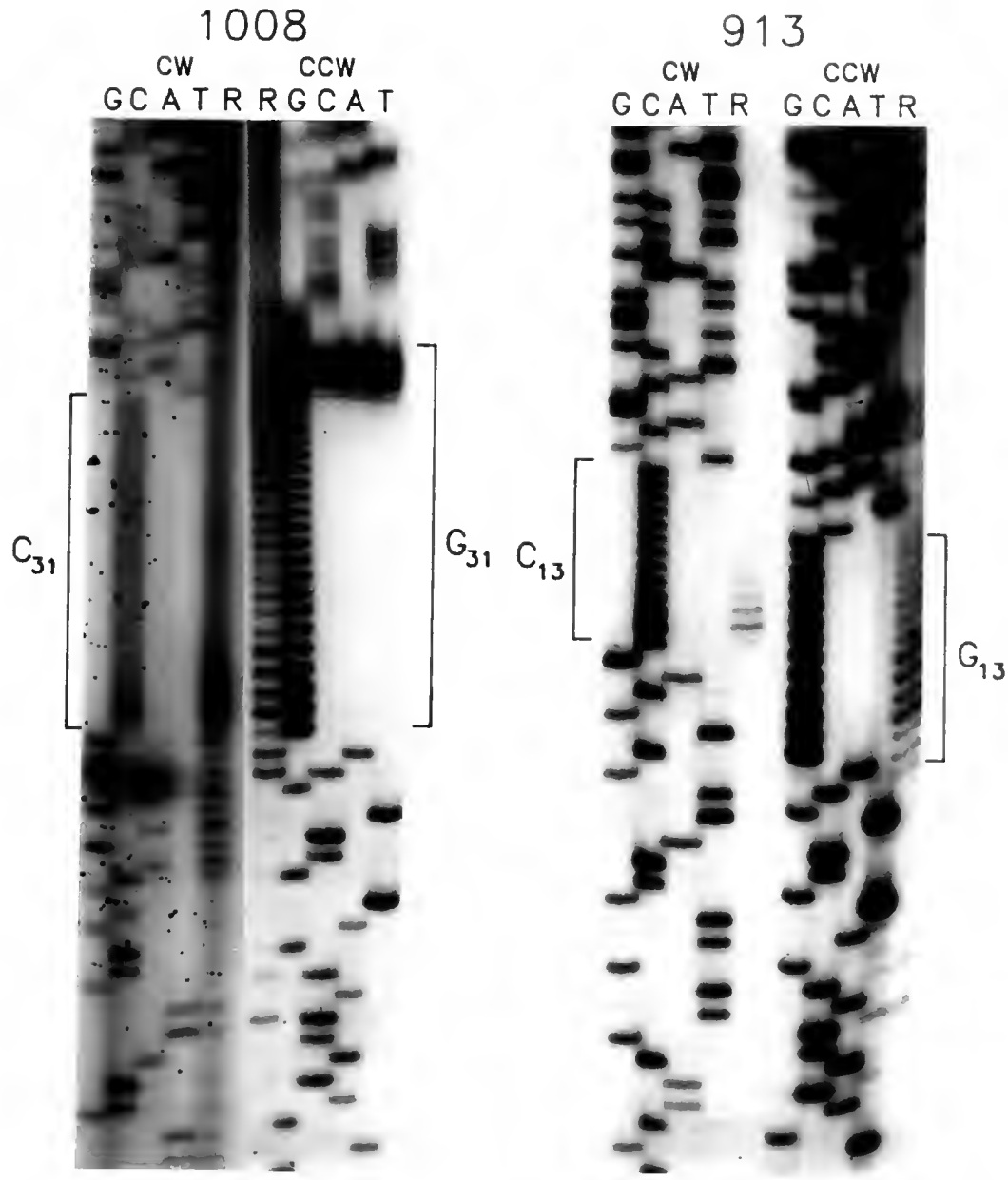
insert/vector junction would produce long single-stranded overhangs at the ends of the fragments. Conceivably, the ends of these fragments could anneal and effectively reduce the observed amount of specifically cleaved product. However, no increase in the amount of specifically cleaved product was observed after the reaction products were phenol extracted and heated, indicating that the cleaved fragments did not have extensive 5′ or 3′ protruding ends capable of hybridization (not shown). Several questions remained, however, concerning the mechanism and pattern of endo R cleavage. For example, it was not known whether endo R cleavage occurred at one or several sites within the recognition sequence or whether there was a difference in the cleavage pattern between substrates that contained different types of recognition signals. Additionally, it was possible that the separate strands of the recognition sequence were cleaved in a dissimilar manner. To address these questions, the sequence at the site of cleavage was determined for several substrates containing various lengths of poly(dG):poly(dC), alternating copolymers of GA, telomeric sequences and the AAV recognition sequences (Figure 5-6). In these studies, the mobility of the primer extension products of endo R fragments were compared with the sequence of the insert and flanking regions of the substrate clone on denaturing acrylamide gels. The sequence of the substrate was determined by the dideoxy sequencing method (Sanger et al., 1977) using the same single-stranded primer used for primer extension (Figures 5-4 A-D, see Chapter II: Materials and Methods). These results were confirmed independently by comparison of the mobility of 5′ and 3′ labeled endo R products with the chemically sequenced plasmid insert (Figure 5-5).

The conclusion from the analysis of the cleavage sites of clones containing homopolymers of poly(dG) was that cleavage could occur throughout the G:C insert (Figure 5-4A, Figure 5-6). A slight preference for cleavage at the 3′ end of both strands was observed in all the homopolymeric clones. This effect was especially pronounced at the 3′ end of the poly(dG) strand in the plasmid pGM1008, which contains the AAV recognition signal 5′ to the G:C insert (Figure 5-6). Cleavage of the G-strand was particularly strong at the G:C/AAV junction and continued into the AAV sequences, while cutting in the C-strand in this clone appears to be more uniform, with only a slight preference for the 3′ end. The site of cleavage in clones that contained only poly(dG):poly(dC) inserts and no AAV sequences (pGM913, pGM1116, Figure 5-6) was also slightly skewed toward the 3′ end of both strands, while cleavage was confined almost entirely to the G:C insert. In substrates that contained the alternating copolymer $(GA)_{11}$ and $(GA)_{38}$ (Figure 5-4B), cutting occurred throughout the polymer insert between the G and A residues on the lower strand and between the C and T residues on the upper strand. In these clones, cleavage was evenly distributed within the CT strand, but more pronounced at the 5′ and 3′ insert/vector junctions in the GA strand.

The cleavage pattern observed with plasmids containing telomeric sequences exhibited more periodicity. In the plasmid pGM1505 (Figure 5-4C), which contains the sequence $(C_{2\text{-}6}T)_5$, cleavage of the CT-strand occurred throughout the insert with a particularly strong cut occurring at regular intervals immediately after the first C residue of every repeat. A preference for cleavage at the 3′ end of the CT-strand was observed, corresponding to the region containing 6 continuous G:C base

Figure 5-4. Sequence at the Site of Endo R Cleavage - Dideoxy Method.

A) Nucleotide sequence determined by the dideoxy method (Sanger, 1977) and primer extension products of endo R digestions for substrates containing: A) poly(dG) inserts; pGM1008 (left) and pGM913 (right), B) alternating copolymers of GA; pGA11 (left) and pGA38 (right), C) telomeric sequences; pGM1505 (left) and pGM1483 (right), and D) a single copy (pGM1228, left) or two tandem inverted copies (pGM1344, right) of the AAV recognition site. The plasmid sequence (GCAT) was determined for each strand of the substrate using either a clockwise (cw) or counterclockwise (ccw) oligonucleotide primer and 1 μg of plasmid. The endo R primer extension products (R) were synthesized from 1 μg of endo R digested substrate with the primer used to generate the corresponding sequence markers (see Chapter II).


1008
CW
CCW
G C A T R R G C A T
C_{31}
G_{31}
913
CW
CCW
G C A T R G C A T R
C_{13}
G_{13}

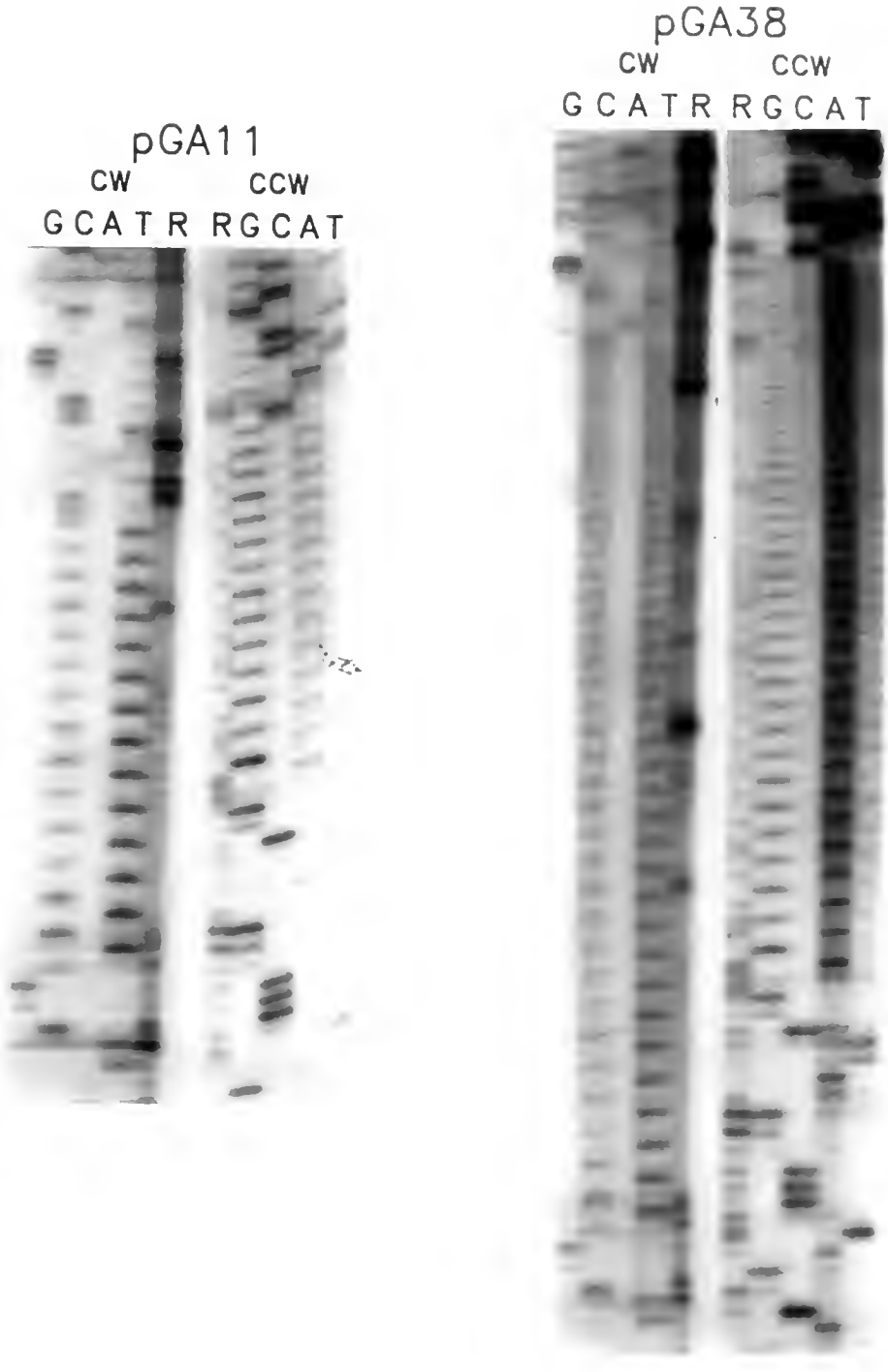


Figure 5-4 (continued), part B

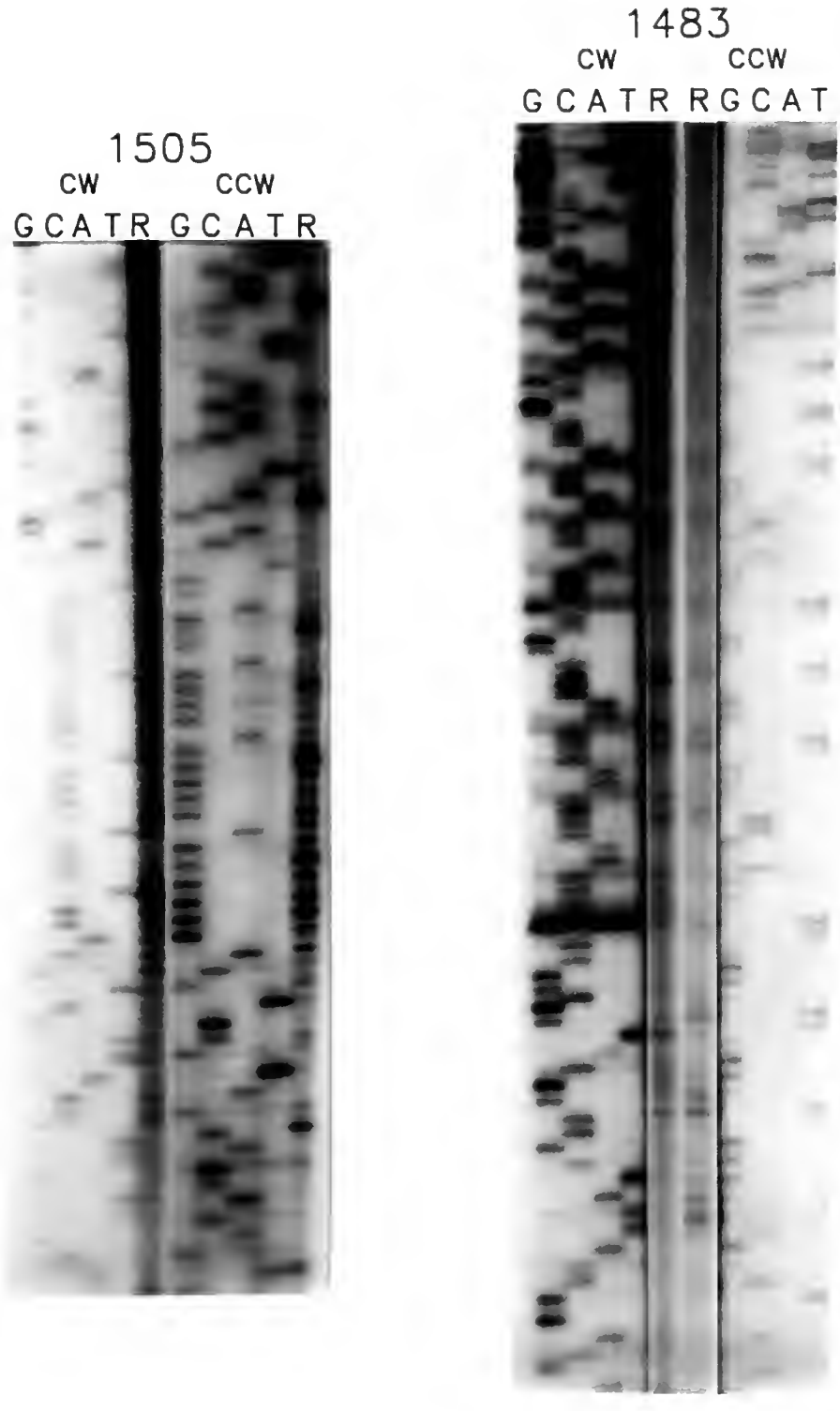


Figure 5-4 (continued), part C

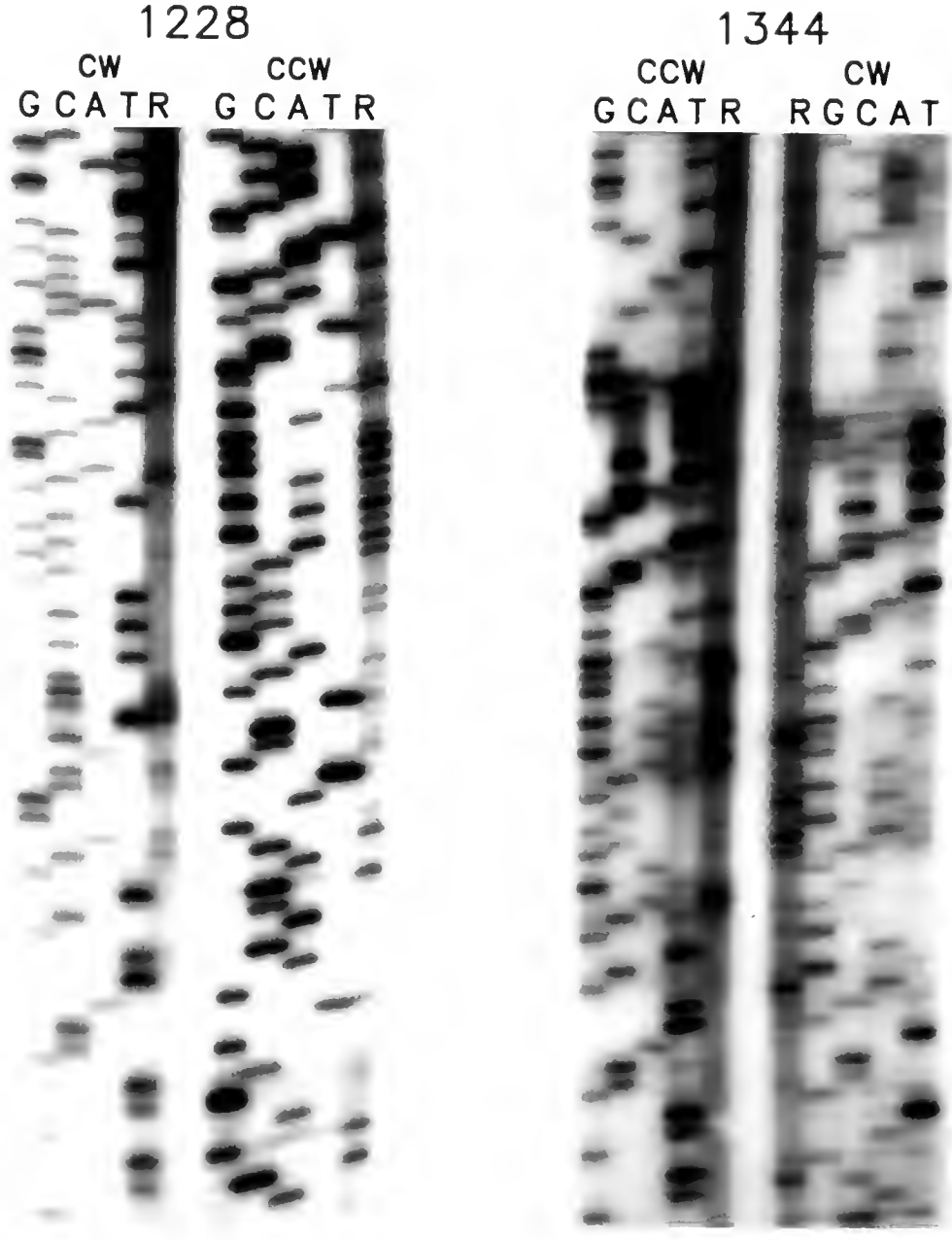


Figure 5-4 (continued), part D

Figure 5-5. Sequence at the Site of Endo R Cleavage - Maxam and Gilbert Method.

Fine mapping of the endo R cleavage products of pGM620D (PstD). The G+A and C+T sequencing reactions were as described (Maxam and Gilbert, 1977). 1 μg of pGM620D labeled at the ScaI site at either the 5' or 3' end was digested with endo R (R) and fractionated on a 8% acrylamide-urea gel in parallel with sequence markers labeled at the same site. The poly$(dG)_{28}$:poly$(dC)_{28}$ tail and the direction of the AAV and pBR322 sequences are indicated in the figure. A detailed description of the sequencing protocol and of PGM620D is included in Chapter II.

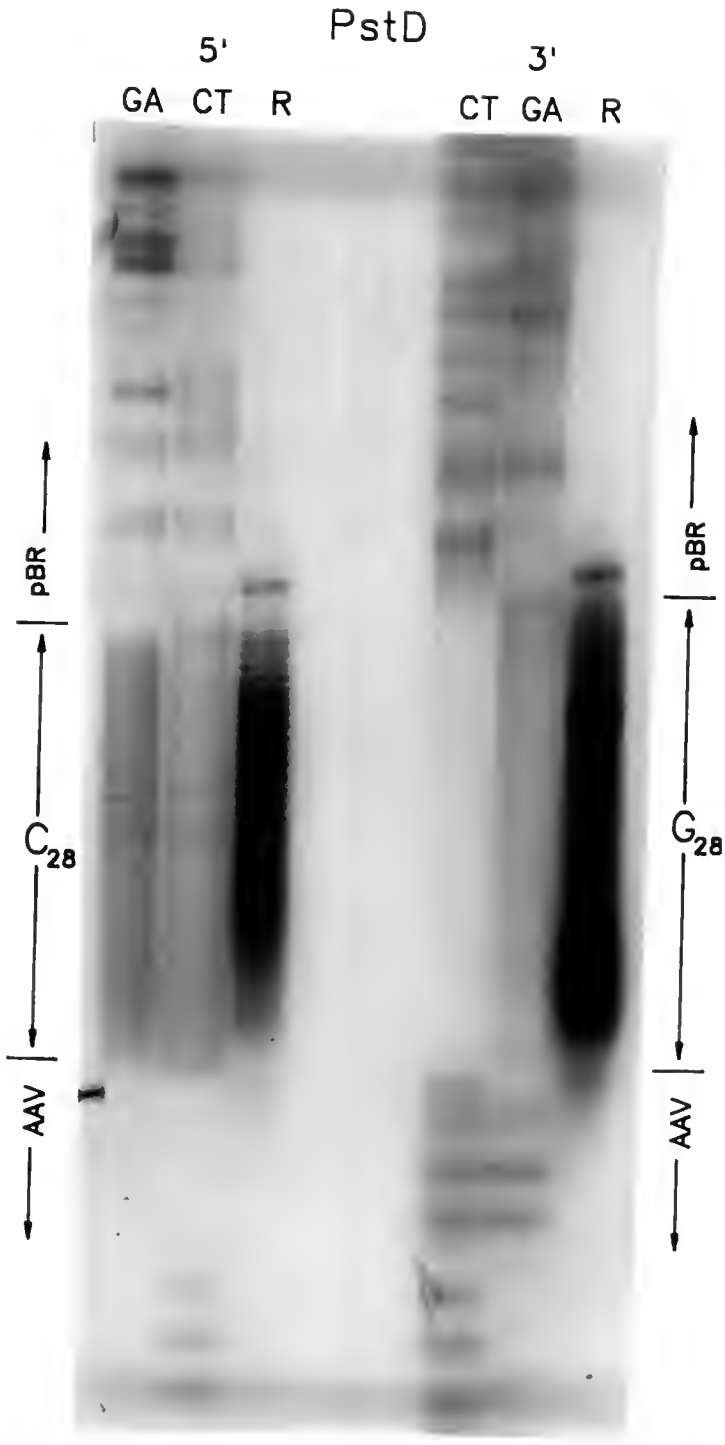
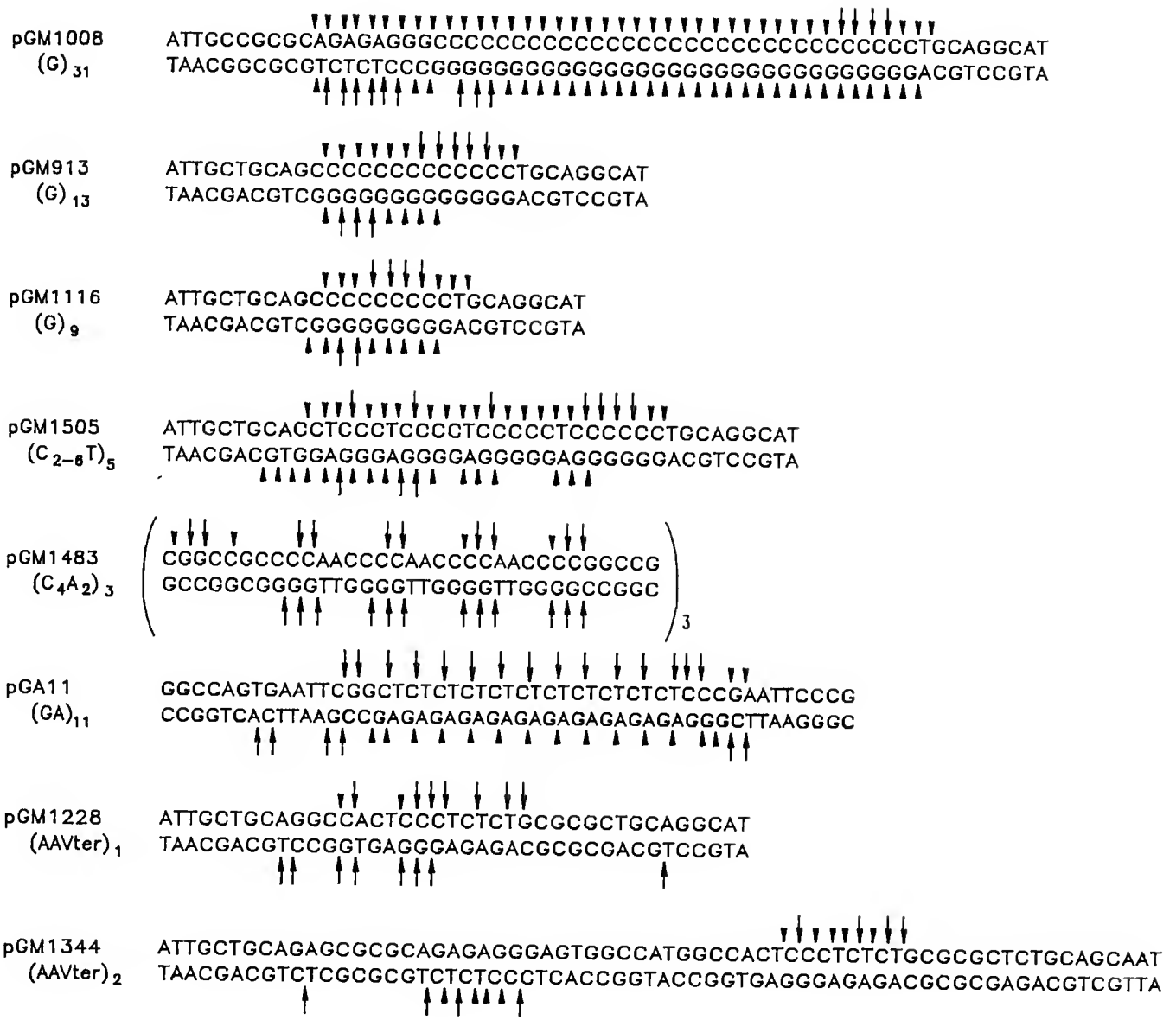

PstD
5'
GA CT R
3'
CT GA R
pBR
C28
AAV
pBR
G28
AAV

Figure 5-6. Sequence at the Site of Endo R Cleavage - Summary of the Cleavage Sites.

The sequence data in Figures 5-4 (A-D) are summarized in this figure. Only the sequence of the oligomer insert and a few bases of the flanking vector regions are shown. The sequences are listed as they appear in the clones, where the top strand runs in a 5′ to 3′ direction. Long arrows represent major cleavage sites and short arrows represent minor sites (usually about 25-50% of the intensity of the major bands). The relative intensities apply only within the same strand and no attempt was made to compare the intensity in opposite strands or between clones (see Table IV for the frequency of cleavage in these substrates).


pGM1008
(G)31
ATTGCCGCGCAGAGAGGGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTGCAGGCAT
TAACGGCGCGTCTCTCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGACGTCCGTA
pGM913
(G)13
ATTGCTGCAGCCCCCCCCCCCCCTGCAGGCAT
TAACGACGTCGGGGGGGGGGGGGACGTCCGTA
pGM1116
(G)9
ATTGCTGCAGCCCCCCCCCTGCAGGCAT
TAACGACGTCGGGGGGGGGACGTCCGTA
pGM1505
(C2–6T)5
ATTGCTGCACCTCCCTCCCCTCCCCCTCCCCCCTGCAGGCAT
TAACGACGTGGAGGGAGGGGAGGGGGAGGGGGGACGTCCGTA
pGM1483
(C4A2)3
CGGCCGCCCCAACCCCAACCCCAACCCCGGCCG
GCCGGCGGGGTTGGGGTTGGGGTTGGGGCCGGC
3
pGA11
(GA)11
GGCCAGTGAATTCGGCTCTCTCTCTCTCTCTCTCTCTCCCGAATTCCCG
CCGGTCACTTAAGCCGAGAGAGAGAGAGAGAGAGAGAGGGCTTAAGGGC
pGM1228
(AAVter)1
ATTGCTGCAGGCCACTCCCTCTCTGCGCGCTGCAGGCAT
TAACGACGTCCGGTGAGGGAGAGACGCGCGACGTCCGTA
pGM1344
(AAVter)2
ATTGCTGCAGAGCGCGCAGAGAGGGAGTGGCCATGGCCACTCCCTCTCTGCGCGCTCTGCAGCAAT
TAACGACGTCTCGCGCGTCTCTCCCTCACCGGTACCGGTGAGGGAGAGACGCGCGAGACGTCGTTA

pairs (Figure 5-6). Curiously, a strong preference for cutting was seen at the 3′ end of the GA strand, even though the longest stretches of G:C base pairs are at the 5′ end of this strand. In contrast, the pattern of cutting observed with the plasmid pGM1483, which contains the sequence $(C_4A_2)_3$, exhibited little preference for either end of the insert (Figure 5-4C). The cleavage pattern for this clone was repetitive and symmetrical throughout the oligonucleotide insert (Figure 5-6). Cleavage appeared to occur in directly opposite positions on each strand immediately before the A residues on the CA strand.

Cleavage of the plasmids pGM1228 and pGM1344, which contain single and inverted double copies of the AAV terminal sequences, respectively, was contained within the C-rich region of the AAV recognition sequence and showed a preference for the 3′ end of each strand (Figure 5-4D, Figure 5-6). This is especially noticeable in pGM1344, where the cut sites are localized exclusively within the C-rich region of each strand and would appear to produce sizable 3′ overhangs.

The results from the sequence studies indicated that cleavage occurred throughout the recognition sequence and frequently extended into the flanking vector sequences. A preference for cleavage at the 3′ end of both strands was observed in most substrates, while cutting in the C-strand had a tendency to be more uniform. In addition, the nucleotide order had no affect on the site of cleavage, where cutting occurred between every base combination. These observations, in conjunction with the observed frequency of cutting with different substrates, suggest that cleavage at the nucleotide level is structural rather than sequence dependent.

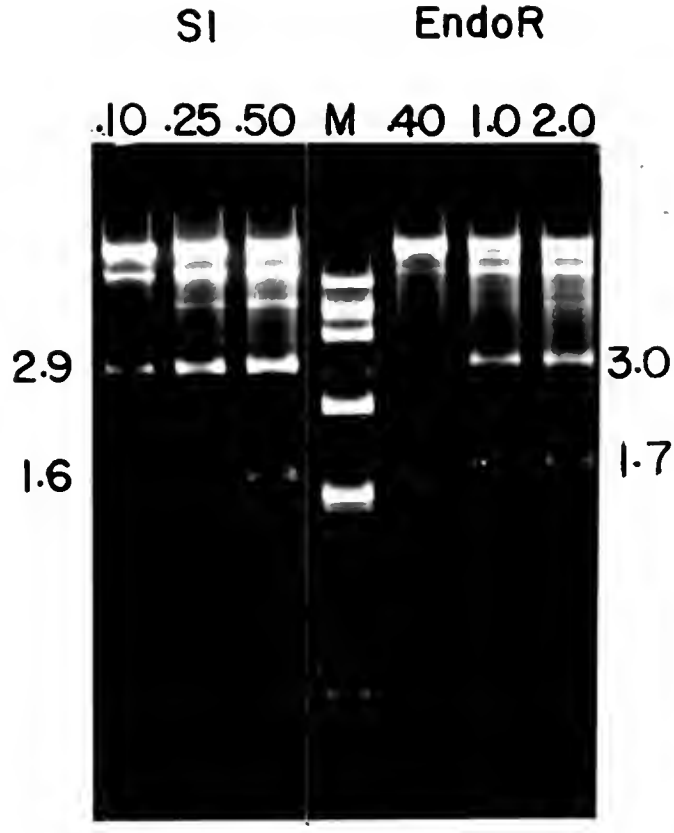
Is Endo R a Site Specific Single-Stranded Nuclease

A number of single-stranded endonucleases will cleave polypurine-polypyrimidine stretches that are present in duplex DNA. The activity has been demonstrated for S1, mung bean, and Neurospora endonucleases as well as for snake venom phosphodiesterase (Evans and Efstradiatis, 1986; Cantor and Efstradiatis, 1984; Schon et al., 1983; Pulleyblank et al., 1985). One important difference between these activities and endo R is that the single-stranded nucleases require supercoiling in substrates that contain short polypurine stretches. In contrast, endo R readily cleaves linear substrates which contain 13 bp of poly(dG) (data not shown). However, there is enough similarity in sequence recognition to suspect that endo R may have properties common to single-stranded endonucleases.

To determine if Endo R was a single-stranded DNA endonuclease, the products of S1 and endo R digestions of pSM620 were compared. S1 produced a cleavage pattern similar to that seen with endo R. However, unlike endo R, S1 did not seem to discriminate between the inboard and outboard ends of the AAV terminal palindrome. It apparently cut equally well at both ends of the terminal repeat, producing a full length pBR322 linear DNA molecule (4.4 kb) and the shorter 1.6 and 2.9 kb AAV fragments (Figure 5-7). The 100 bp deletion resulting from the loss of the AAV terminal sequences is especially noticeable when comparing the 1.7 kb endo R products with the 1.6 kb fragments produced by S1. It is not known whether the loss of the AAV terminal sequences was due to selective degradation of the palindrome or to specific cleavage at both the inboard and outboard recognition sites. Similar experiments with mung bean endonuclease did not produce discrete bands (not shown).

Figure 5-7. S1 and Endo R Nuclease Digestions of pSM620.

One microgram (0.2 pmol) of pSM620 DNA was titrated with S1 nuclease or fraction IV endo R in a 50 μl reaction volume under standard conditions for each enzyme (see Chapter II: Materials and Methods). Numbers at the top indicate the number of units of each enzyme used.


S1
EndoR
.10 .25 .50 M .40 1.0 2.0
2.9
1.6
3.0
1.7

S1 nuclease typically releases acid soluble oligo- and mono-nucleotides as products of the reaction (Vogt, 1973). After normalizing endo R and S1 activities on duplex substrates, the activities on single-stranded ^{3}H-labeled E.coli DNA were compared by measuring the release of acid soluble radioactivity (Figure 5-8). When equivalent amounts of S1 and endo R (as measured on duplex DNA) were incubated with ^{3}H-labeled single-stranded E.Coli DNA, S1 converted 70% of the DNA to acid soluble product within 30 minutes, while endo R produced no acid soluble products after 2 hours. The results were similar regardless of whether the reactions were carried out at pH 6.0 for both enzymes, or whether each enzyme was assayed at its own pH optimum (pH 4.5 for S1 and pH 7.5 for endo R). On the basis of this assay, it was estimated that endo R was approximately 50-100 fold less active on single-stranded DNA than S1.

The ability of endo R and S1 to degrade single-stranded circular DNA was also compared. Under standard reaction conditions, single-stranded circular ϕX and M13 DNA are degraded to oligonucleotides of 200-400 base pairs in length (Figure 5-9). When assayed by gel mobility, endo R cut single-stranded circular DNA at a rate 20 fold lower than S1 (not shown). This was also true in reactions which contained standard amounts of plasmid substrate in addition to ϕX DNA (not shown). In these experiments, 50% of the plasmid substrate was cleaved specifically, while activity on single-stranded circles was similar to that observed without plasmid substrate. It was concluded from these experiments that endo R was clearly not a typical single-stranded nuclease. However, it was not clear whether endo R activity on

single-stranded circles was an intrinsic property of endo R or the result of a residual contaminating endonuclease.

Figure 5-8. S1 and Endo R Activity on Single-Stranded DNA.

Five micrograms of single-stranded, uniformly ^{3}H-labeled E. coli DNA was incubated with S1 nuclease (1.25 units) or endo R (5.0 units) in a 250 μl reaction at 37^{o}C under standard reaction conditions for each enzyme (see Materials and Methods). Equivalent amounts of double-stranded cleavage activity for both enzymes were used for each reaction (See Figure 5-7). At the indicated times, 25 μl portions of each reaction mixture were acid precipitated and counted. The percent (%) acid-precipitable counts remaining at each time point is plotted. Closed circles, endo R; open circles, S1.

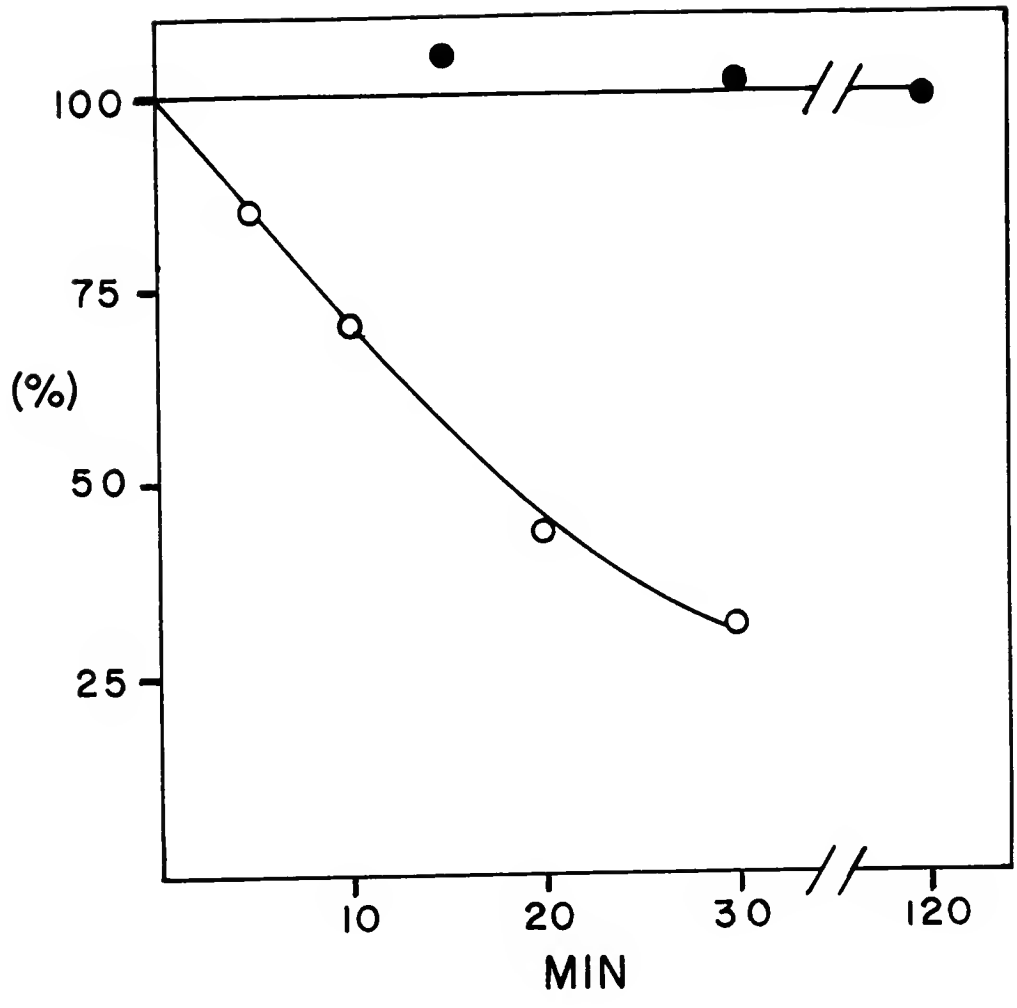
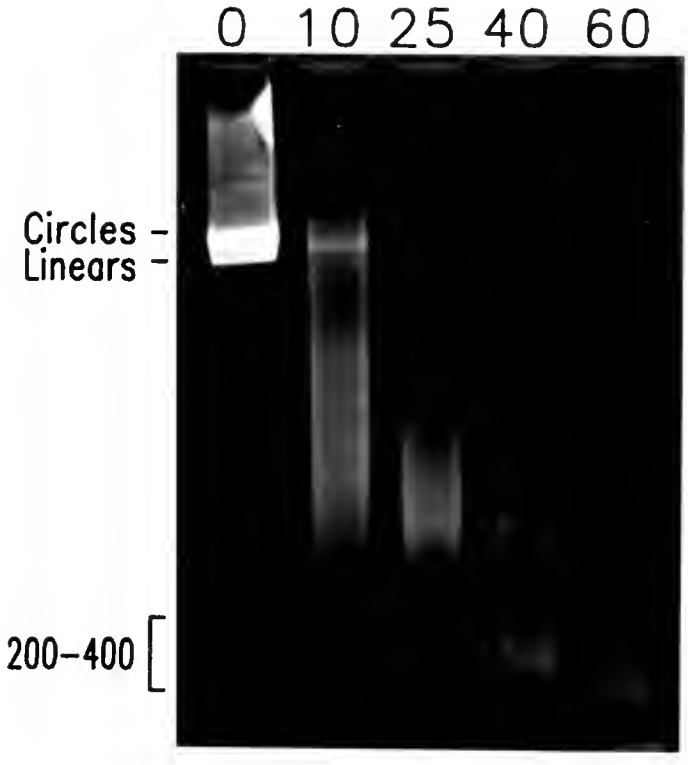

100
75
(%)
50
25
10
20
30
120
MIN

Figure 5-9. Endo R Activity on Single-Stranded Circular DNA.

A standard reaction mixture of 125 μl contained 5 μg of ϕX single-stranded circular DNA and 2.5 units of endo R. At the indicated times, 25 μl portions were removed, the reaction was terminated with the addition of 0.1% SDS and 20 mM EDTA, and the reaction products were fractionated on a 1.4% agarose gel.


0 10 25 40 60
Circles
Linears
200–400

CHAPTER VI
PROPERTIES OF ENDO R

Reaction Conditions for Double-Stranded Cleavage

Purified endo R requires Mg^{2+} cations at 5-10 mM for optimal cleavage activity (Table V). Calcium ions (Ca^{2+}) at 0.1-1.0 mM can substitute for Mg^{2+}, but only about half of the activity is achieved under these conditions. No activity is observed with Zn^{2+} alone at any concentration, and both Ca^{2+} and Zn^{2+} are inhibitory at higher concentrations in the presence of Mg^{2+}. When Mn^{2+} is substituted for Mg^{2+}, a different specificity of cleavage is observed at concentrations below 0.1 mM, while greater concentrations result in non-specific degradation of the substrate.

Endo R is extremely salt sensitive (Figure 6-1). A 50% inhibition was observed with NaCl or KCl at concentrations of 10 mM. Because of this sensitivity, fractions emerging from columns eluted with salt gradients were dialyzed before being assayed.

Endo R is active over a wide pH range, from pH 6 to pH 9, in Tris-HCl, sodium acetate or sodium phosphate buffers (not shown). The pI of endo R was estimated to be between pH 6.5 and 7.0. This was determined by observing the binding characteristics of the enzyme to DEAE cellulose at different pHs (see Chapter II: Materials and Methods). No increase in the rate of the reaction was observed with temperatures above 37°C, while temperatures below 37°C resulted in a decreased rate of cleavage.

TABLE V

DIVALENT METAL CATION
REQUIREMENTS FOR ENDO R[a]

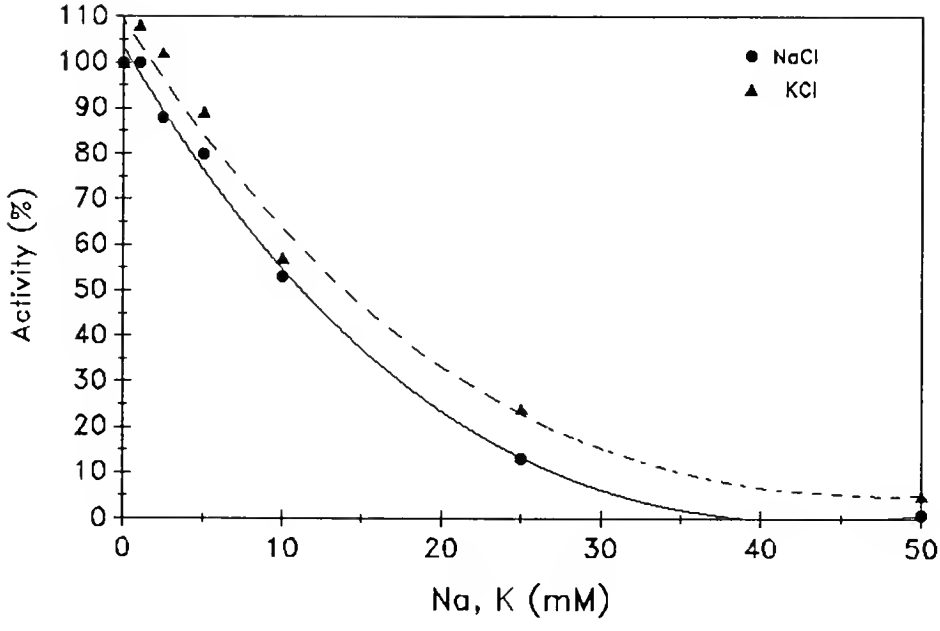
$MgCl_2$ (mM)	$MnCl_2$[a] (mM)	$CaCl_2$ (mM)	$ZnCl_2$ (mM)	% Double Stranded Cleavage
0	-	-	-	0
1.0	-	-	-	50
5.0	-	-	-	100
10.0	-	-	-	100
15.0	-	-	-	80
-	0	-	-	0
-	0.1	-	-	100
-	>0.5	-	-	0
-	-	0.1	-	50
-	-	0.5	-	50
-	-	1.0	-	50
-	-	5.0	-	5
-	-	10.0	-	<1
5.0	-	0.1-1.0	-	100
5.0	-	5.0	-	50
5.0	-	10.0	-	40
	-	-	0-10.0	0
5.0	-	-	0.1	40
5.0	-	-	>0.1	0
	-	-		

[a]*Bst*EII linearized pSM620 (0.2 pmol) was incubated with 1 unit of fraction V endo R under standard reaction conditions with the amount of divalent metal cation indicated in the table. The reaction products were fractionated on 1.4% agarose gels, stained with ethidium bromide and photographed. The amount of double-stranded cleavage was determined from densitometer tracings of the 3.0 and 1.7 kbp endo R/*Bst*EII products as described in the legend to figure 6-1.

[b]Concentrations of $MnCl_2$ >0.1 mM result in nonspecific degradation of substrate DNA, while concentrations between 0.01 and 0.10 mM produce a different cleavage specificity (see Figure 6-3).

Figure 6-1. Sodium and Potassium Inhibition of Endo R.

Standard reactions containing 0.2 pmol of BstEII linearized PSM620 were incubated with 1 unit of fraction V endo R with increasing concentrations of NaCl or KCl for 1 hour at 37°C. The products of the reaction were fractionated on a 1.4% agarose gel, stained with ethidium bromide and photographed with Polaroid Type 55 film, which produces a negative and a positive print. The intensity of the 3.0 kbp product was measured from the negative using the LKB ultrascan XL densitometer and the LKB gel scan program, version 1.0. The percent activity is plotted against the concentration of Na and K chloride in the reaction. The amount of cleavage at 1 mM NaCl (the salt contribution from the extracts) is assigned 100% activity.


110
100
90
80
70
60
50
40
30
20
10
0
Activity (%)
0
10
20
30
40
50
Na, K (mM)
NaCl
KCl

Reducing agents such as 2-mercaptoethanol and dithiothreitol had little or no effect on activity (not shown).

Substrate concentrations between 20 and 40 μg/ml were optimal for double-stranded cleavage. Purified enzyme exhibited no preference for form I or form III plasmid substrates, and both were cleaved at equal rates. When the appearance of endo R digested products was compared over time, the amount of product formation was identical for both types of input plasmid (Figure 6-2). This suggests that the conversion of form I substrate to form III was not the rate limiting step in the production of specifically cleaved products and that supercoiled or torsionally stressed substrates were not required for endo R specific cleavage.

Taking these factors in account, standard reaction mixtures of 25 μl were incubated at 37°C for 1 hour and contained 20 mM Tris-HCl, pH 7.5; 5 mM $MgCl_2$; 1 mM DTT, 0.2 pmol form I plasmid substrate and 1 unit of endo R. One unit of endo R activity is defined as the amount of protein that is required to cleave 50% of pGM620D substrate under standard reaction conditions.

Alternate Cleavage Specificity in the Presence of Mn^{2+} Ions

When manganese is used as metal ion cofactor, complete degradation of the substrate occurs with standard levels of enzyme and a different specificity is observed with low concentration of Mn^{2+} or enzyme. In Figure 6-3, peak fractions emerging from a Bio-Gel 1.5M column (fraction IV) were assayed in an identical series of reactions using either Mg^{2+} of Mn^{2+} as the cofactor. Comparison of the activity profiles revealed that the different specificities were coincidental. High concentrations of enzyme resulted in high levels of specific cleavage with Mg^{2+}, but

Figure 6-2. Cleavage of Form I and Form III Plasmid Substrate.

Two and one-half micrograms of supercoiled (Fml) or BstEII linearized (Fm3) pSM620 substrate was incubated with 2.5 units of fraction V endo R in reaction volumes of 125 μl under standard reaction conditions. At the indicated times, 25 μl portions were removed and treated with proteinase K stop solution as described in the legend to Figure 3-2. After phenol extraction and ethanol precipitation, the endo R products of the form I substrate were digested with BstEII for 1 hour at 60^{o}C. Both the BstEII digested and undigested reaction products were electrophoresed on a 1.4% agarose gel.

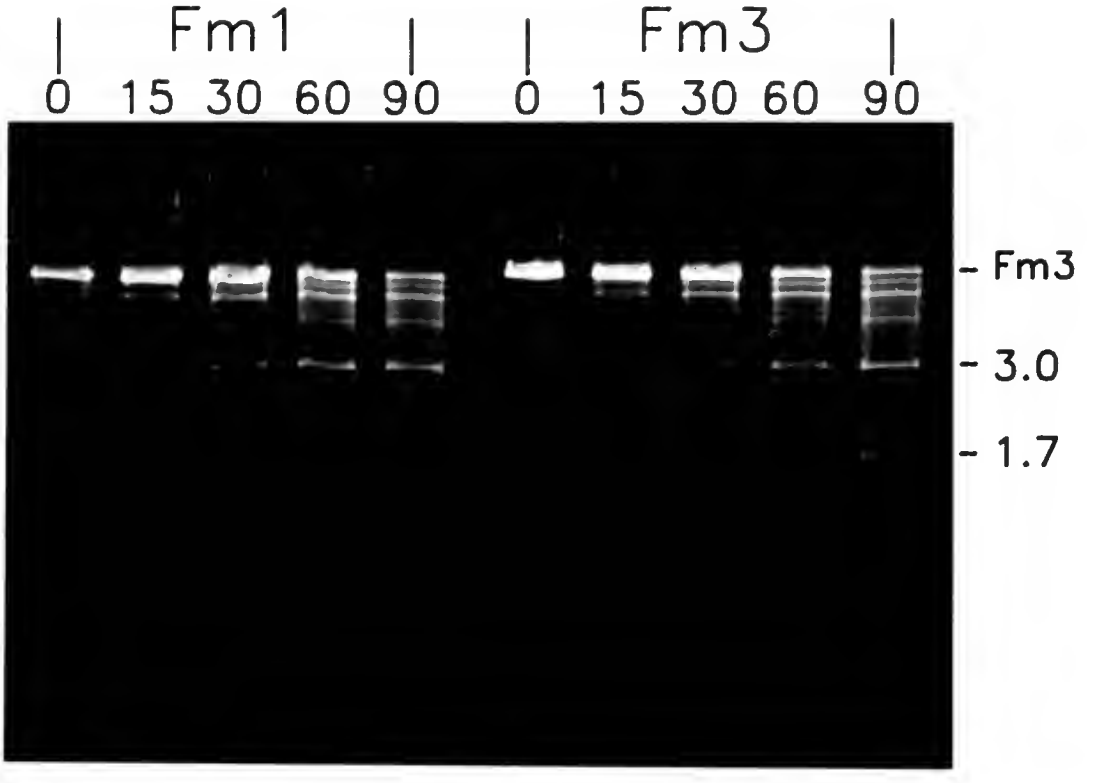
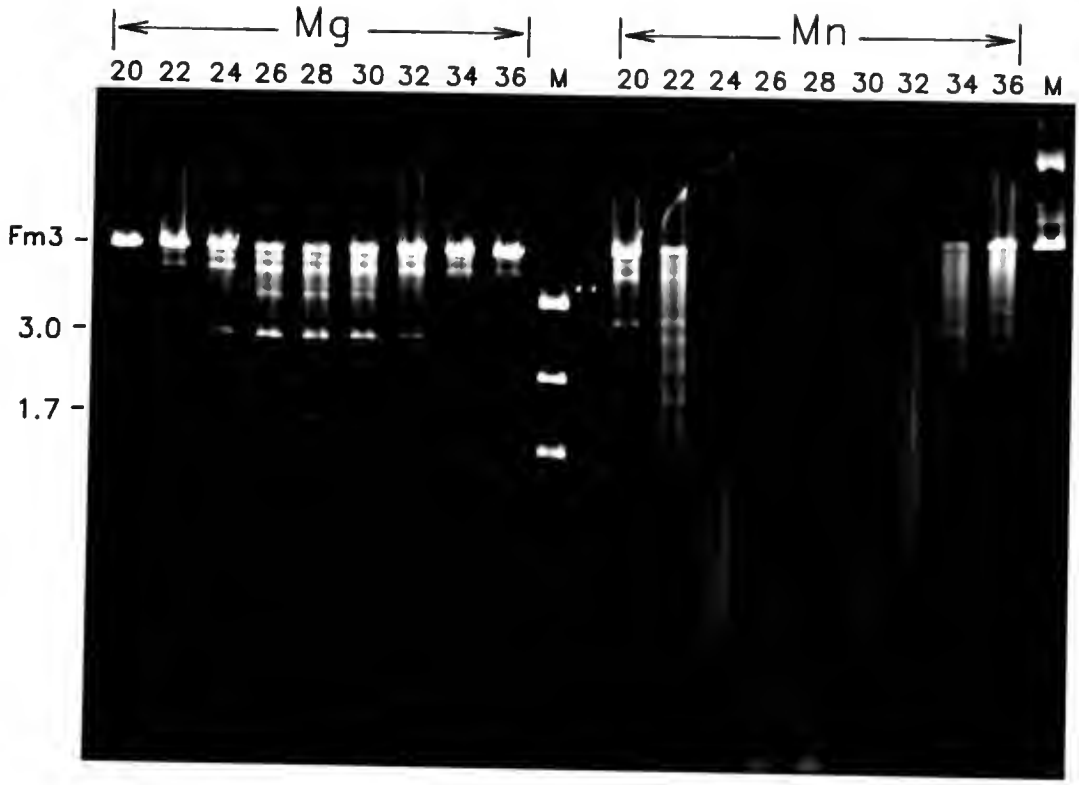

Fm1
Fm3
0 15 30 60 90 0 15 30 60 90
- Fm3
- 3.0
- 1.7

Figure 6-3. Specificity of Endo R with Mg^{2+} and Mn^{2+}.

Standard reaction mixtures of 25 μl containing 0.5 μg of pSM620 substrate, 5.0 mM $MgCl_2$ or $MnCl_2$, and 4 μl of the indicated fraction from a BioGel A1.5M column (Fraction IV endo R), were incubated at 37°C for 1 hour. The samples were phenol extracted, ethanol precipitated, digested with BstEII, and fractionated on a 1.4% agarose gel.


Mg
Mn
20 22 24 26 28 30 32 34 36 M
20 22 24 26 28 30 32 34 36 M
Fm3 -
3.0 -
1.7 -

degraded the substrate entirely in the presence of Mn^{2+}. Conversely, low enzyme concentrations present in fractions around the perimeter of the peak, resulted in an altered specificity with Mn^{2+} as cofactor, and a reduced endo R cleavage with Mg^{2+}. These results suggest that the different specificities reside on the same protein and are intrinsic properties of endo R observed under different conditions.

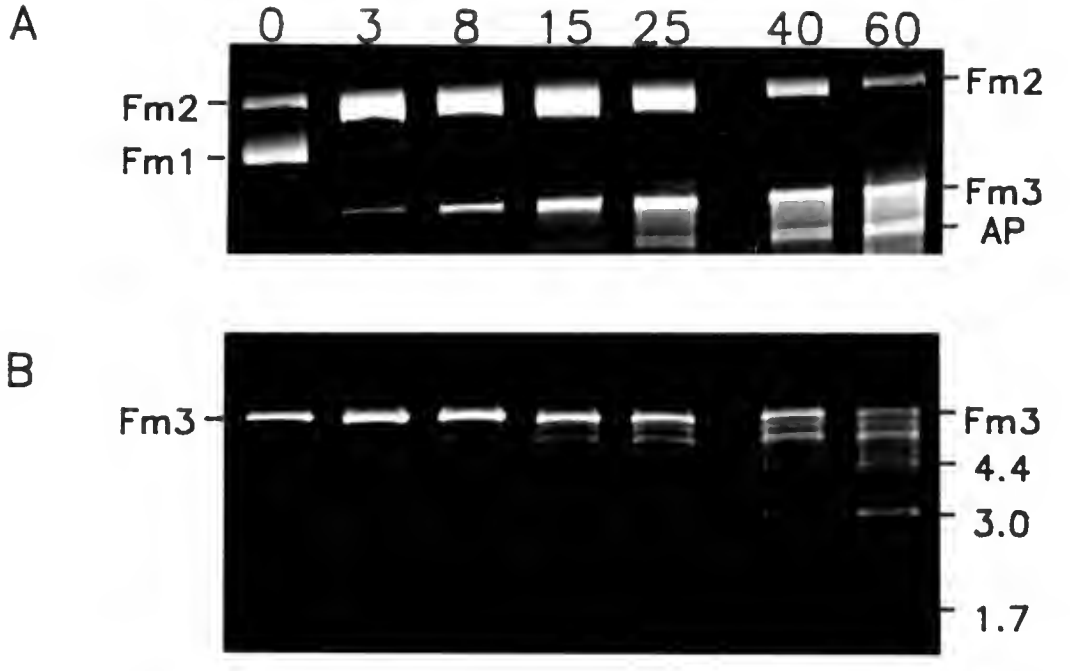
Endo R Specific Cleavage is of a Double-Stranded Nature

Analysis of the endo R products at several time points during the reaction before digestion with BstEII (Figure 6-4A) revealed that the supercoiled input plasmid was quickly converted to the nicked circular plasmid form. Linear pSM620 molecules, which are formed from one double-stranded break, were produced as early as three minutes into the reaction, while the production of AAV and pBR322 linears, which require two double-stranded breaks, were detected at 15 minutes. Comparison of the BstEII digested products (Figure 6-4B) with undigested samples revealed that specific cleavage at the AAV/pBR322 junctions paralleled the production of form III plasmid and AAV and pBR monomers. The accumulation of linear molecules and specifically cleaved products correlated well with the decrease in form II molecules. This relationship is displayed graphically in figure 6-4C.

Previous results indicated that endo R was equally active on supercoiled and linear forms of input plasmid and that the rate of formation of specifically cleaved double-stranded products paralleled the rate of formation of linear molecules. It was possible, however, that the production of linear molecules was preceded by a nicked circular intermediate as an obligatory step, and that the form II

Figure 6-4. Cleavage Reaction Time Course.

Two picamoles of form I pSM620 was incubated with 20 units of fraction V enzyme at 37°C in a reaction volume of 0.5 ml. 50 μl portions were removed at the indicated times and the reactions were terminated with the addition of an equal volume of stop solution (1 mg/ml proteinase K, 20 mM EDTA, 1.0% SDS), phenol extracted, and ethanol precipitated. One half of each time point was digested with BstEII and both the uncut (A) and BstEII digested (B) samples were fractionated on a 1.4% agarose gel. Fm3 = linear plasmid, A and P are linear monomer AAV and pBR322 molecules, respectively. C) The conversion of the undigested plasmid forms (in A, F1 and F2) to linear endo R products (F3) are plotted as a function of time. The appearance of the 3.0 kb (3K) fragment from BstEII digestion of endo R products was used as a measure of the level of specific cleavage of the supercoiled substrate. The level of cleavage at 60 minutes was assigned 100%.


A
0
3
8
15
25
40
60
Fm2
Fm1
Fm2
Fm3
AP
B
Fm3
Fm3
4.4
3.0
1.7

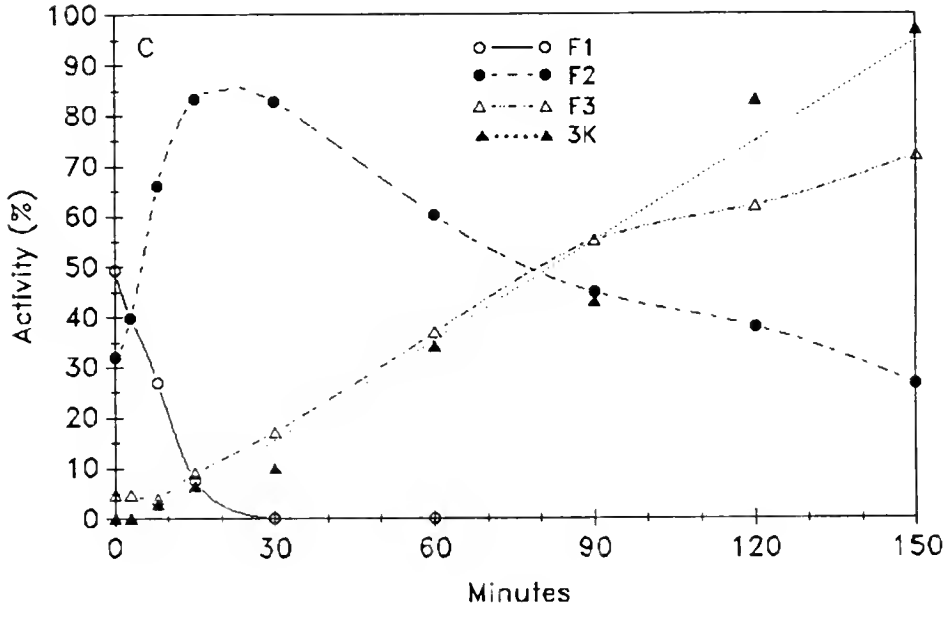
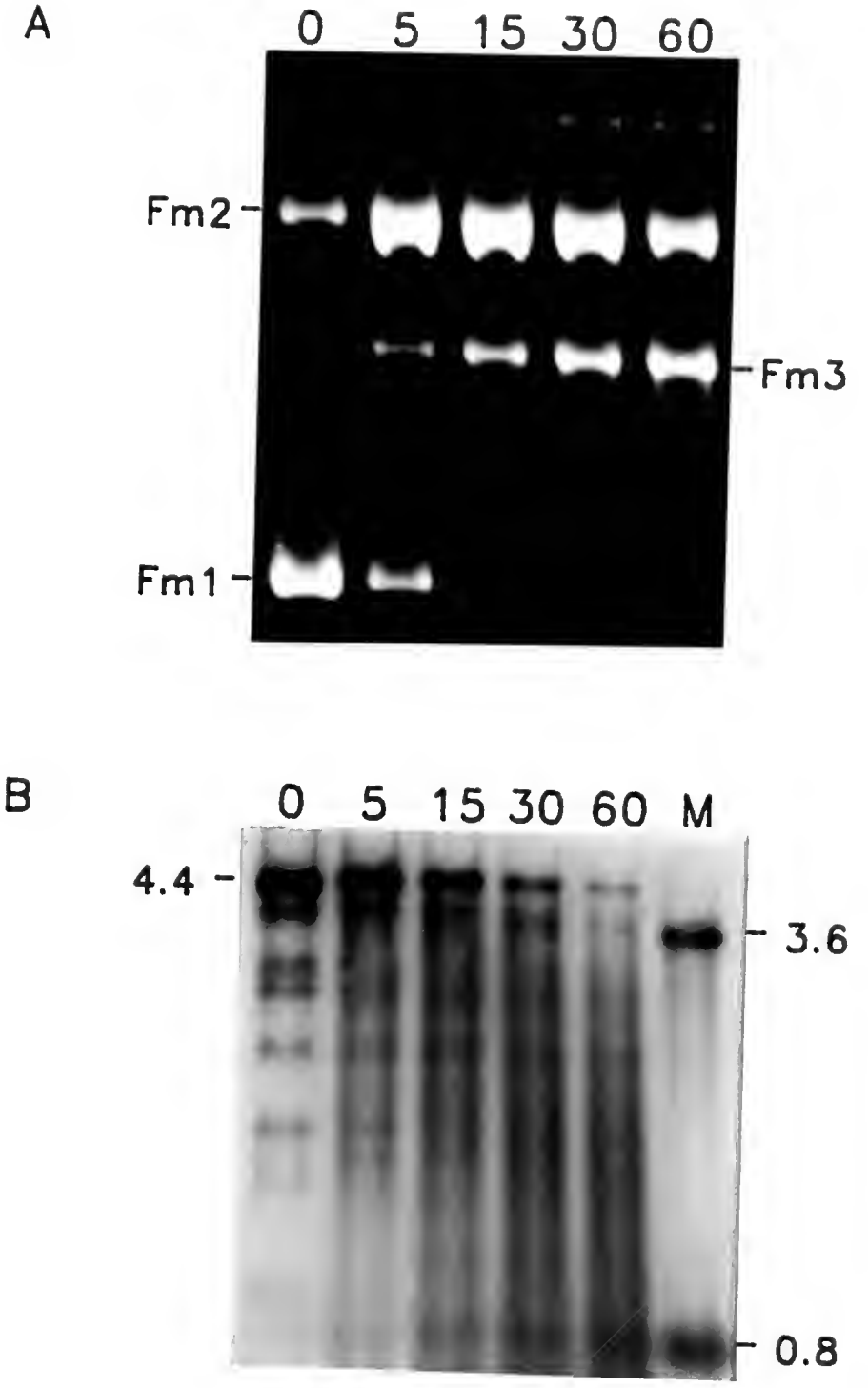


Figure 6-4 (continued), part C

Figure 6-5. Single-Stranded Specificity of Endo R.

A reaction mixture of 500 μl containing 1.0 pmol of pGM1008 in standard reaction buffer was incubated with 20 units of endo R (fraction V) at 37^{o}C. At the indicated times, 50 μl aliquots were removed and treated as described for figure 3-2. One half of each time point was digested with EcoRI, denatured with the addition of 0.1 volume of 2 N NaOH, and fractionated on a 1.5% denaturing agarose gel (B), as described in Chapter II: Materials and Methods. Time points not digested with EcoRI were fractionated on a 1.0% non-denaturing agarose gel (A). The 0.8 and 3.6 kilobase fragments are produced by either single or double-stranded cleavage by endo R and further digestion with EcoRI .


A
0
5
15
30
60
Fm2
Fm3
Fm1
B
0
5
15
30
60
M
4.4
3.6
0.8

molecules were produced by site specific single-stranded cleavage of the input plasmid. Digestion of a specifically cleaved form II intermediate with a one-cut restriction enzyme would yield nicked form III molecules that would be indistinguishable from unreacted material on native agarose gels, but would produce discrete bands on a denaturing agarose gel. In addition, nicked circular products were observed early in the reaction and would result in the early accumulation of specifically cleaved products. These could be easily distinguished from products formed from double-stranded cuts which occurred relatively late in the reaction. To detect the presence of site specific single-stranded breaks, the plasmid pGM1008 was incubated with endo R and the products of the reaction were analyzed at several time points by fractionation on both neutral and denaturing agarose gels (Figures 6-5 A and B, respectively). As can be seen from these experiments, the production of the 0.8 kb fragment produced from the EcoRI digestion of endo R products (Figure 6-5B) proceeded slowly and corresponded closely to the accumulation of form III in undigested products (Figure 6-5A). This indicates that the formation of nicked circular DNA very early in the reaction was the result of random nicking of the substrate and that specific double-stranded cleavage occurs at a much slower rate.

Effect of Nucleotides and Polynucleotides on the Cleavage Activity

The addition of substances that stimulate or inhibit activity can often reveal important aspects of enzymatic mechanisms. With a nuclease such as endo R, it was logical to look at the effect that the addition of nucleotides and nucleotide polymers would have on specific cleavage activity.

The addition of a mixture of the 4 deoxyribonucleoside triphosphates (dNTP) at concentrations between 0 and 5 mM had no observable effect on the endo R cleavage reaction (Table VI). However, when added separately, dGTP and dATP stimulated cleavage 2-3 fold at concentrations between 0.5 and 1.0 mM. In contrast, the addition of dCTP or dTTP had no observable effect on activity. Conversely, a mixture of the 4 ribonucleoside triphosphates (rNTP) was slightly inhibitory at 5mM. CTP and UTP did not affect activity when added separately, while GTP of ATP completely inhibited double stranded cleavage at concentrations greater than 5mM. Curiously, the addition of ATP at concentrations greater than 1.0 mM produced a pronounced inhibition (Figure 6-6). Moreover, the addition of an ATP generating system (creatine phosphate and creatine phosphokinase), in the presence of 1.0 mM ATP, completely inhibited double-stranded cleavage. However, endo R has shown no affinity for GTP or ATP-linked agarose resins in the presence of Mg^{2+} ions, and no other indication of binding between endo R and ATP or GTP could be demonstrated. In crude preparations of endo R, the addition of duplex calf thymus DNA and ATP resulted in a stimulation of specific cleavage. Subsequent experiments with more purified preparations of enzyme have shown that this effect was probably due to inhibition of interfering activities (i.e. contaminating nucleases) rather than to a direct stimulation of endo R. In addition, endo R (fractions III, IV, or V) did not hydrolyze γ^{3}H-ATP under standard cleavage reaction conditions. No ATPase activity could be detected in these fractions with or without the addition of substrate DNA (not shown).

TABLE VI

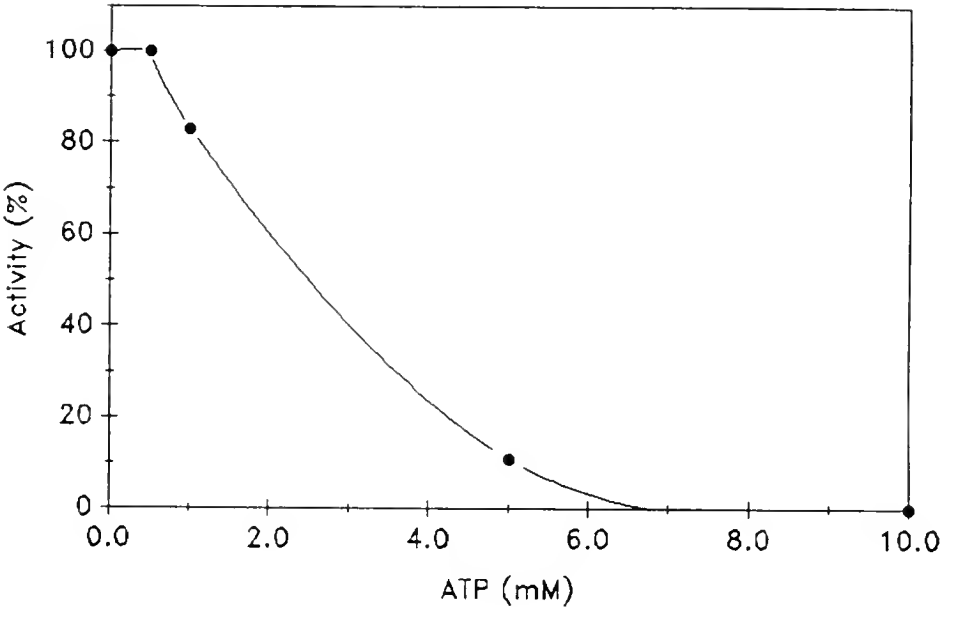
NUCLEOTIDE INHIBITION AND STIMULATION[a]

	Concentration μg/ml	Activity %
dNTP	0-5.0	100
dGTP	0-0.1	100
	0.5	150
	1.0	200
	5.0	90
dCTP	0-5.0	100
dATP	0-0.05	100
	0.1	150
	0.5	200
	1.0	150
	5.0	10
dTTP	0-0.1	100
	0.5-5.0	125
rNTP	0-1.0	100
	1.0	90
	5.0	80
GTP	0-1.0	100
	5.0	10
CTP	0-5.0	100
ATP	0-0.5	100
	1.0	25
	5.0	<1
UTP	0-1.0	100
	5.0	10

[a]Standard reaction mixtures (25 μl) containing 0.2 pmol of BstEII digested pSM620 was incubated with 1 unit of endo R at 37°C for 1 hour with the indicated concentration of nucleotide. The reaction products were fractionated on 1.4% agarose gels and the relative activity was estimated from photographs and densitometer tracings as described for Figure 6-1. dNTP = all four deoxyribonucleotide triphosphates; rNTP = all four ribonucleotide triphosphates.

Figure 6-6. ATP Inhibition of Endo R Activity.

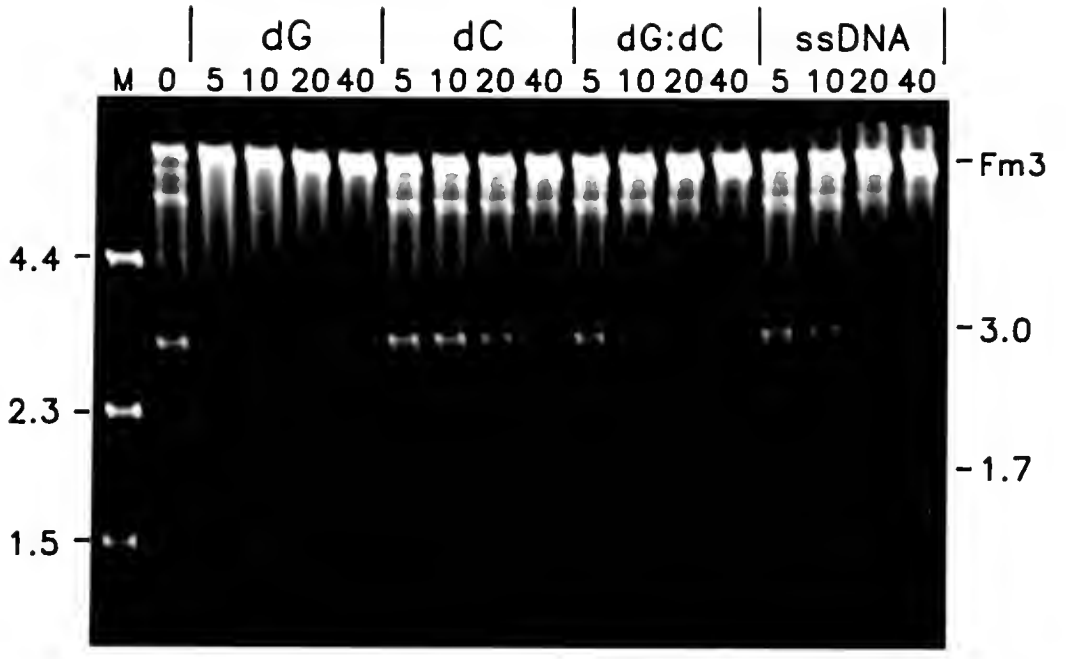
ATP at concentrations of 0, 0.1, 0.5, 1.0, 5.0 and 10.0 mM was added to standard endo R cleavage reactions (25 μl) containing 0.5 μg of BstEII digested pSM620 substrate. The quantity of product formed was determined as described for Figure 6-1 using the LKB ultrascan densitometer.


100
80
60
40
20
0
Activity (%)
0.0
2.0
4.0
6.0
8.0
10.0
ATP (mM)

The results of assays in which homopolymers were added to the reaction indicated that endo R had a strong affinity for polymers of single-stranded poly(dG) over other DNA forms. In these experiments, standard endo R reactions containing 20 μg/ml pSM620 substrate, were titrated with single-stranded $poly(dG)_{12}$, single-stranded $poly(dC)_{12}$, native E.coli DNA, heat-denatured E.coli DNA or with double-stranded $poly(dG)_{12}$:$poly(dC)_{12}$ (Figure 6-7A). The results clearly showed that single-stranded poly(dG) DNA was the most potent inhibitor of endo R activity, and completely inhibited cleavage at all concentrations tested. The quantity of poly(dG) required for complete inhibition is 10-100 fold less than the concentrations required of the other competitor DNAs, and at least four times less than the number of moles of nucleotides in the substrate. Double-stranded poly(dG):poly(dC), total cellular RNA (not shown) and native (not shown) or denatured E.coli DNA inhibited the cleavage reaction at approximately equal concentrations. Approximately 40 μg/ml of each were required to achieve 90% inhibition (Figure 6-7B), which corresponds to 2 moles of nucleotide inhibitor per mole of nucleotide substrate. The addition of tRNA or polymers of brominated poly(GC), which forms a left-handed Z-DNA conformation, had no effect on cleavage activity at any of the concentrations tested (0-40 μg/ml). The affinity of endo R for single-stranded poly(dG) was confirmed by chromatographic results, where only minimal amounts of activity bound to single-stranded DNA agarose, while all of the cleavage activity bound tightly to single-stranded poly(dG) agarose columns of comparable capacity.

Figure 6-7. Polynucleotide Competition of Double-Stranded Cleavage Activity.

A) Standard reactions mixtures (25 μl) containing 20 μg/ml BstEII digested pSM620, 1 unit of fraction V endo R and 0, 5, 10, 20 or 40 μg/ml DNA competitor were incubated for 1 hour at 37°C. The products were electrophoresed on a 1.4% agarose gel and the level of endo R activity was measured as described for Figure 6-1. dG = single-stranded $poly(dG)_{12}$, dC = single-stranded $poly(dC)_{12}$, GC = double-stranded $poly(dG)_{12}$:$poly(dC)_{12}$ and SS = heat denatured E.coli DNA. B) The cleavage activity is plotted against the amount of competitor DNA added to the reaction. The activity profile for double-stranded E.coli DNA and single-stranded E.coli DNA were virtually identical and only the data for the single-stranded competitor is shown here.


dG
dC
dG:dC
ssDNA
M 0 5 10 20 40 5 10 20 40 5 10 20 40 5 10 20 40
4.4
2.3
1.5
Fm3
3.0
1.7

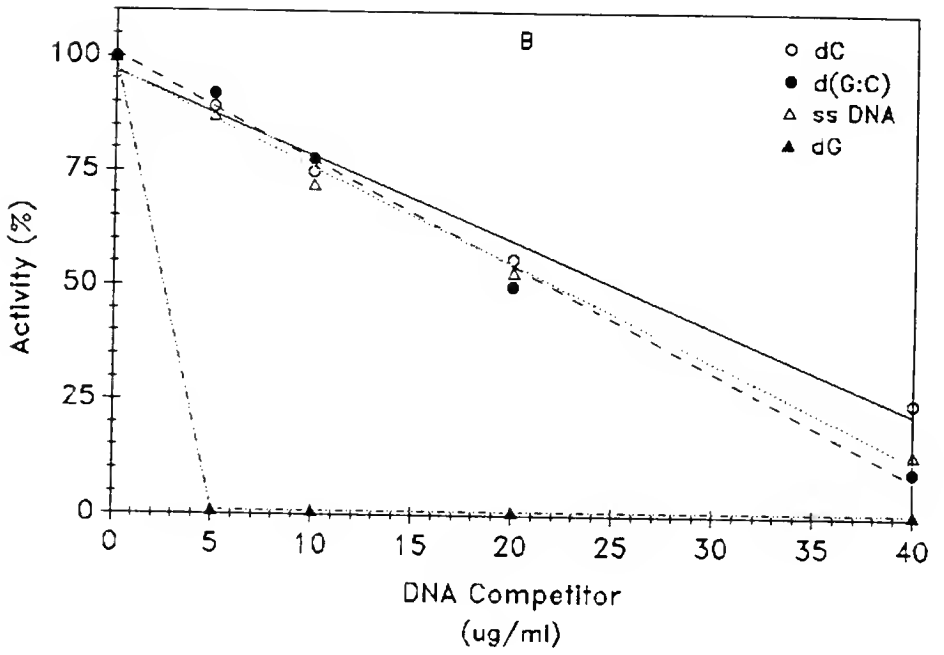
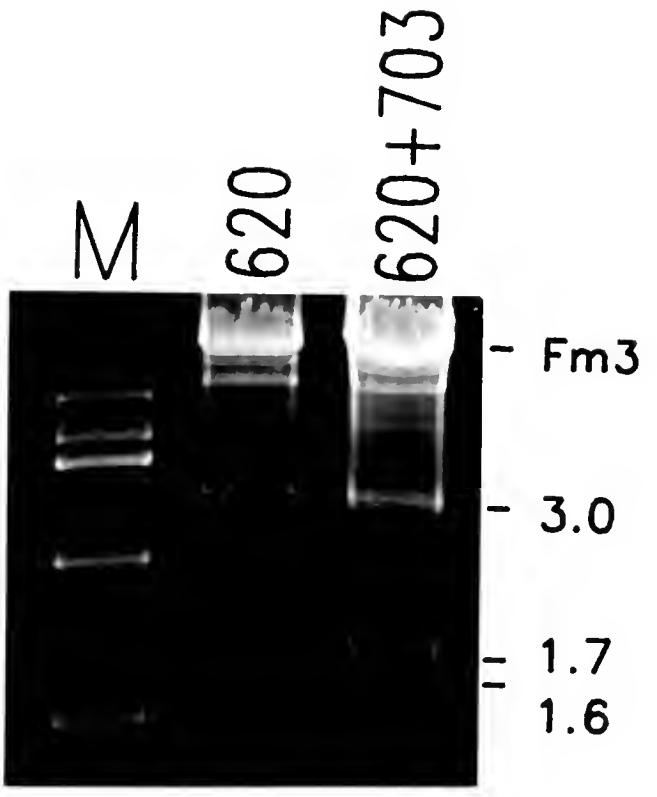


Figure 6-7 (continued), part B

Figure 6-8. Stoichiometry of the Endo R Cleavage Reaction.

A reaction mixture containing 1.0 pmol of pSM620 in 100 μl was incubated with 1 unit of fraction V endo R under standard reaction conditions. At 30 minutes, the reaction was split into two portions, the reaction terminated in one portion with the addition of an equal volume of 1 mg/ml proteinase K, 20 mM EDTA, 1.0% SDS and incubation at 37^{o}C for 1 hour. 0.5 pmols of pSM703 was added to the remaining portion, incubation at 37^{o}C was continued for an additional 30 minutes, and the reaction was terminated as above. After phenol extraction and ethanol precipitation, both aliquots were digested with BstEII and fractionated on a 1.4% agarose gel. The 1.7 kb fragment is indicative of the amount of pSM620 cleavage, while the 1.6 kb band indicates the amount of pSM703 cleavage.


M
620
620+703
Fm3
3.0
1.7
1.6

Are Stoichiometric Amounts of Enzyme Required for Cleavage

To determine if endo R was required in stoichiometric amounts for cleavage, two different substrates which could be differentiated on the basis of the size of the cleavage products, were added sequentially to the reaction mixture (Figure 6-8). When a 10 fold excess of pSM620 was incubated with endo R and digested with BstEII, only a small fraction of the substrate had reacted at 30 minutes, indicating that the reaction was saturated with substrate. The addition of an equal amount of pSM703 to the reaction, followed by incubation for an additional 30 minutes, resulted in the continued cleavage of both substrates at a maximum rate. Cleavage of the pSM620 substrate can be differentiated from pSM703 cleavage by virtue of the 100 base pair deletions in the AAV terminal sequences in the pSM703 clone. BstEII digestion of endo R products produces a 1.7 kb fragment from cleavage at the left pSM620 junction, which is easily distinguished from the 1.6 kb fragment generated by cleavage at the left pSM703 junction (refer to Figure 4-3). These results indicated a distributive or non-stoichiometric mode of double-stranded cleavage by endo R.

Characterization of the Ends of the Reaction Products

Knowledge of the position of the phosphomonoester and hydroxyl groups on the products of nucleolytic digestion is useful for comparison with other nucleases. The positions of these groups were determined independently by the ligation of endo R products and by 5′ and 3′ labeling experiments. The plasmid pGM1008 was used as the substrate for endo R ligation studies. PGM1008 is cleaved by fraction V endo R at a frequency equal to that of pSM620 (Table IV). The initial products of

endo R digestion of pGM1008 consisted of approximately equal portions of form II and form III DNA and included a background smear, which was presumably the combined result of endo R cleavage at minor sites and nonspecific degradation of the substrate (Figure 6-9, lane R). When these products were treated with T4 DNA ligase, greater than 50% of the linear molecules could be ligated to higher molecular weight forms (Figure 6-9, lane RL). No increase in the yield of higher molecular weight products was observed with continued ligation or after treatment of endo R products with the Klenow fragment to produce blunt ends prior to ligation (not shown). These results indicated that a majority of endo R products contained ends with 5′ phosphates and 3′ hydroxyls and were, therefore, substrates for DNA ligase. However, a significant amount of the form III product was not ligatable. This may reflect a heterogeneous nature of the ends produced by endo R, where only a fraction of the reaction products were ligatable forms that did not contain gaps or overhangs in the hybridized molecule. Generating blunt ends prior to ligation may produce more ligatable forms, but also reduces the overall efficiency of ligation. It was also possible, however, that incomplete ligation reflected the presence of contaminating nucleolytic or phosphatase activities in the enzyme extracts.

The 5′ position of the phosphomonoester group was further verified by 5′ end labeling experiments (Table VII). In these studies, endo R treated samples that had been dephosphorylated with calf intestine alkaline phosphatase (CIAP) prior to end labeling, incorporated 50 fold more radioactivity than the untreated control samples. However, a significant amount of incorporation was detected in endo R products that

TABLE VII

END LABELING OF ENDO R PRODUCTS[a]

Reaction Conditions	Incorporation (cpm x 10^{-3})
No Endo R, No CIAP	8
No Endo R, + CIAP	6
+ Endo R, No CIAP	24
+ Endo R, + CIAP	110
pBR322 x EcoRI	6

[a]Reaction mixtures (50 μl) containing 0.4 pmols of the plasmid pGM1008 were incubated with 2 units of fraction V endo R under standard reaction conditions for 1 hour at 37°C. The reactions were split into two equal portions and 1 portion was treated with calf intestine alkaline phosphatase (CIAP). Both the phosphatased and phosphorylated samples and two identical control samples, which had not been treated with endo R, were 5'-end labeled with polynucleotide kinase and γ^{32}P-ATP as described in Chapter II. The kinasing reactions (25 μl) were terminated with the addition of 2 μl of 0.25 M EDTA, a 5 μl aliquot of each was precipitated in 10% TCA, 0.2 mg/ml BSA and the acid precipitable radioactivity of each fraction was determined.

were not dephosphorylated, corresponding to about 5 fold less than the incorporation observed with the CIAP-treated products. It was possible that this incorporation was due to a kinase exchange reaction. However, it may also reflect the presence of an additional nucleolytic or phosphatase activity in the extracts.

To determine if endo R extracts contained contaminating nucleolytic and phosphatase activities, a portion of the products of the end-labeling reaction was digested with EcoRI, and both the undigested (Figure 6-10A) and EcoRI digested (Figure 6-10B) samples were fractionated on agarose gels. The results in Figure 6-10A show that most of the label in the phosphatased, endo R treated samples was incorporated into form III and form II plasmid DNA. When these products were further digested with EcoRI, the majority of the label was retained either in the specifically cleaved fragments or in unreacted substrate (Figure 6-10B, right panel). Samples not dephosphorylated before end labeling displayed a similar pattern of incorporation, but at a much reduced level. It is unlikely that the same nucleolytic activity would produce two types of ends (Laskowski, 1985). Therefore, the radioactivity incorporated into endo R products not treated with CIAP may be the result of a kinase exchange reaction or may reflect the presence of an associated phosphatase activity in the enzyme extracts. Finally, there is no evidence to support the presence of a contaminating endonuclease, since the pattern of incorporation was the same regardless of whether the samples were dephosphorylated prior to end labeling. However, the results do not exclude the possibility that a small amount of contaminating exonucleolytic activity is present in the extracts. Overall, the results from these studies indicate that the majority of

Figure 6-9. Ligation of Endo R Products.

A standard reaction mixture of 100 μl containing 0.4 pmol of form I pSM620 substrate was incubated with 4 units of fraction V endo R for 1 hour at 37°C. The reaction was phenol extracted and ethanol precipitated and the dry DNA pellet was redissolved in 100 μl of T4 DNA ligase buffer (50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 1 mM DTT and 1 mM ATP). The mixture was divided into 2 equal portions and 400 units of T4 DNA ligase were added to one of the portions. The ligation reaction was incubated at 15°C for 24 hours and both the ligated (RL) and unligated (R) endo R products were fractionated on a 1% agarose gel. The position of nicked circular (Fm2) and linear (Fm3) plasmid DNA is indicated next to the marker lane.

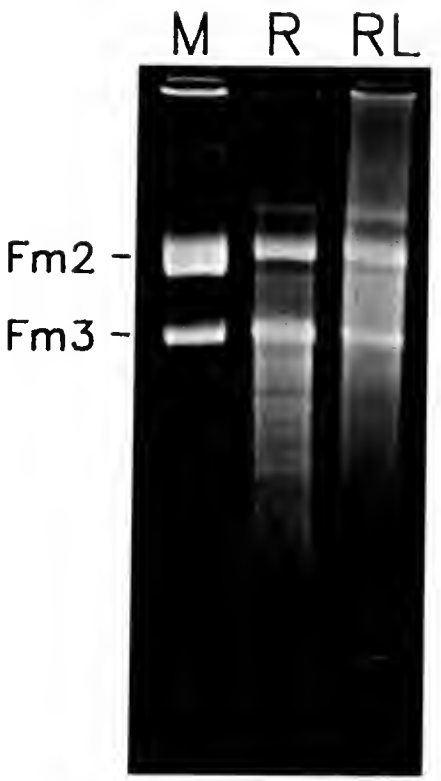
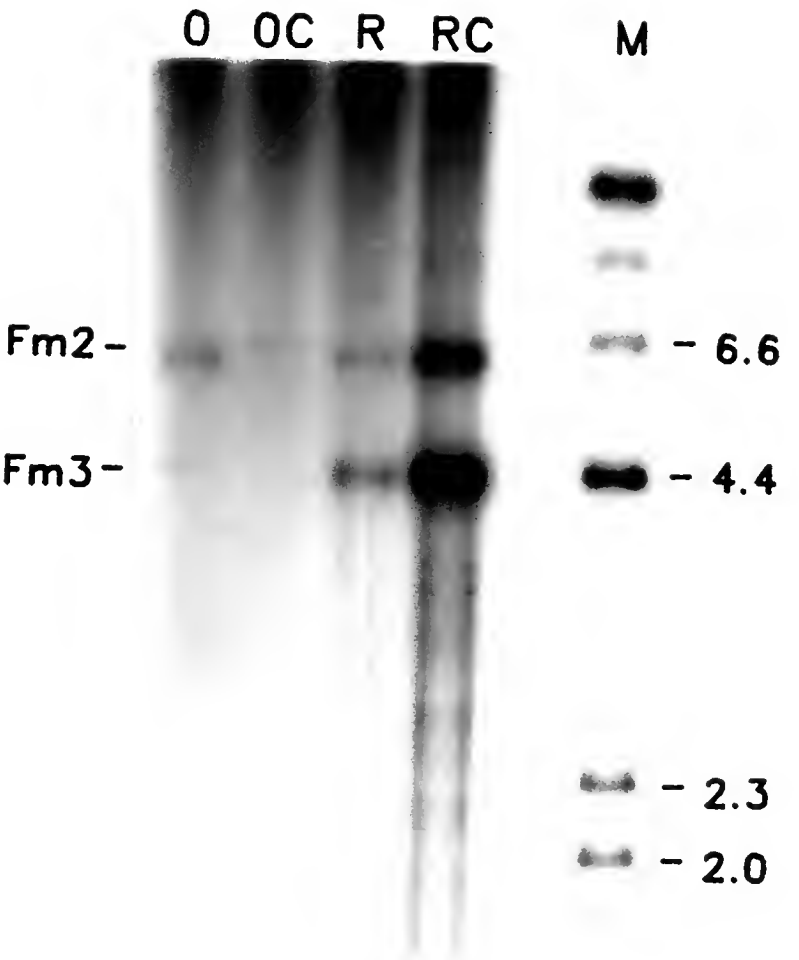

M
R
RL
Fm2
Fm3

Figure 6-10. 5′ End Labeling of Endo R Products.

PGM1008 substrate (0.4 pmol) was incubated with 2 units of fraction V endo R in a standard reaction of 50 μl. A second identical control reaction was not treated with endo R. The reaction mixtures were phenol extracted and ethanol precipitated and the pellets were redissolved in 25 μl of 0.1 M Tris-HCl, pH 8.5, and 0.1% SDS. One half of each mixture was dephosphorylated with the addition of 1 unit of calf intestine alkaline phosphatase (CIAP) and incubation at room temperature for 2 hours. All samples were phenol and chloroform extracted, ethanol precipitated and the pellets were redissolved in 25 μl polynucleotide kinase buffer (50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 1 mM DTT) containing 10 μCi of γ^{32}P-ATP and 4 units kinase. After incubation at 37°C for 30 minutes, a 5 μl aliquot of each sample was precipitated in 10% TCA, 0.2 mg/ml BSA and the amount of acid precipitable counts was determined. O = no endo R; OC = no endo R plus CIAP; R = endo R treated; RC = endo R treated plus CIAP. A) Autoradiograph of undigested products B) Ethidium stained gel (left panel) and autoradiograph (right panel) of EcoRI digested 5′-end labeled samples.


0
OC
R
RC
M
Fm2
Fm3
6.6
4.4
2.3
2.0

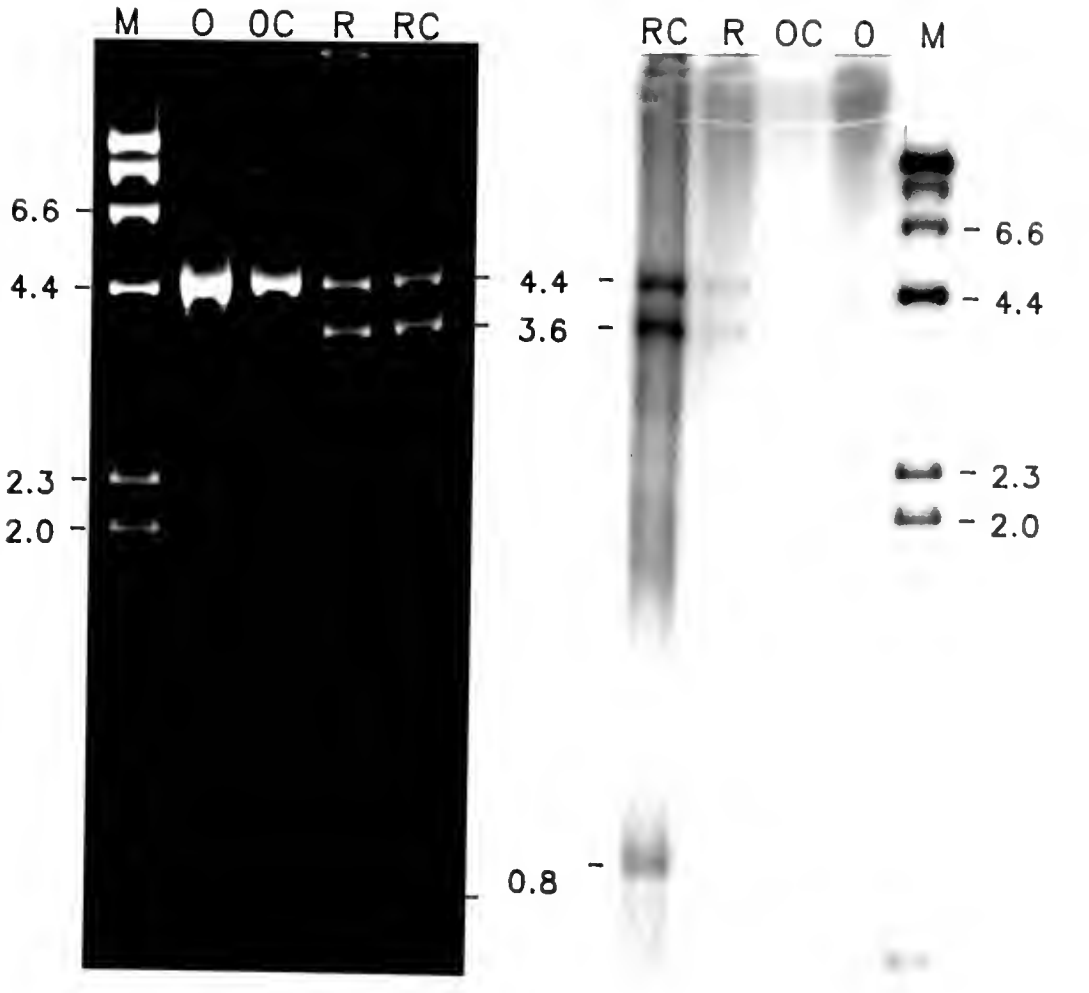


Figure 6-10 (continued), part B

ends of the fragments produced by endo R specific cleavage contain 5' phosphate and 3' hydroxyl groups.

Is Endo R a Human Topoisomerase

The properties of endo R share a few similarities with those of mammalian topoisomerases. The most intriguing is the capacity of both type I and type II topoisomerases to introduce transient breaks in DNA (Liu et al., 1983). It was natural to speculate then, that endo R cleavage is actually due to interruption of the relaxation process of a supercoiled substrate. This is especially true of type II topoisomerases, where double-stranded cleavage has been functionally separated from the rejoining step with the use of specific inhibitors (Chen et al., 1984; Nelson et al., 1984; Yang et al., 1985).

Established assays for the detection of topoisomerase activity (Liu et al., 1980, 1981) were employed to determine if endo R contained associated topoisomerase activity. The available evidence overwhelmingly indicates that endo R does not contain an associated topo activity. In contrast to type I or type II topoisomerases, topoisomers have never been observed in agarose gels when form I plasmid DNA was incubated with endo R. In addition, antibodies to both types of human topoisomerases did not significantly inhibit double-stranded cleavage (not shown). Finally, although endo R did decatenate kinetoplast DNA (not shown), the products of the reaction were linear molecules and not the monomer circles characteristic of type II topoisomerase activity (Liu et al., 1980).

CHAPTER VII
DISCUSSION AND CONCLUSIONS

Purification of Endo R

Endo R is an enzyme of cellular origin. However, specific cleavage activity is enhanced approximately 5 fold in nuclear extracts prepared from Ad2 infected cells grown in the presence of hydroxyurea. Hydroxyurea is a potent reversible inhibitor of ribonucleotide reductase and inhibits DNA replication by blocking the production of deoxyribonucleotides from ribonucleotide precursors (Timson, 1975). Thus, the observed stimulation of activity is presumably the result of Ad2 early gene expression, which does not require DNA replication (Green et al., 1971), and is therefore, unaffected by the addition of hydroxyurea. It is not clear, however, whether the stimulation results from enzyme modification or from a direct enhancement of endo R gene expression.

There remains a significant difference between the molecular weights determined for native and denatured protein. The gel filtration and sedimentation data indicate a native molecular weight of approximately 115-120 kd, while the results from SDS gel analysis indicate that endo R is a slightly smaller protein of approximately 100,000 dal. This type of discrepancy is not uncommon, and could be explained by postulating an asymmetric, nonspherical shape for native endo R. It should be noted however, that endo R has not been purified to homogeneity. Therefore, it is possible that the minor peptides in the enzyme preparation contribute significantly to cleavage activity.

Excision of AAV Sequences In Vitro

When crude nuclear extracts were assayed for specific cleavage activity on AAV plasmid substrates, only 1% of the starting material was cleaved at either AAV/vector junction and the products could be visualized only by Southern (1975) hybridization. Further purification of endo R has simplified the gel assay and greatly increased the yield of specifically cleaved product. This has allowed the detailed study of AAV excision *in vitro*, and has led to a better understanding of the characteristics of endo R nucleolytic activity.

The wild type AAV plasmid, pSM620, contains two major endo R sites at either AAV/vector junction. The major product of the reaction was a linear plasmid molecule produced by independent cleavage at one or the other junction. Therefore, the fragments observed in the agarose gel assay were the result of a partial digestion with endo R followed by complete digestion with *Bst*EII. On occasion, however, both AAV/vector junctions were cleaved simultaneously, accounting for approximately 3% of the product and yielding linear AAV and pBR322 molecules. The site of cleavage in pSM620 was localized to within 25 bp of the AAV/vector junction by hybridizing the reaction products to either AAV of pBR322 specific probes. In addition, it was demonstrated that interaction between the two AAV terminal repeats was not required and that either AAV/vector junction could be cleaved independently of the other. Subclones that contained either the left or right AAV/pBR322 junction were cleaved with equal efficiency when compared to cleavage of the parental clone.

Endo R cleavage of mutant AAV clones in which the AAV/vector junctions had been sequenced, aided in the identification of a

recognition signal for the enzyme. The conclusion from these experiments was that endo R recognized polypurine/polypyrimidine sequences rich in G:C base pairs. This was verified further by the identification of minor fragments produced from cleavage within internal AAV sequences. A computer search using a consensus sequence derived from these experiments, reliably predicted the presence of most of the fragments observed from endo R cleavage of the AAV substrates.

Is Endo R responsible for the Rescue of AAV Sequences In Vivo

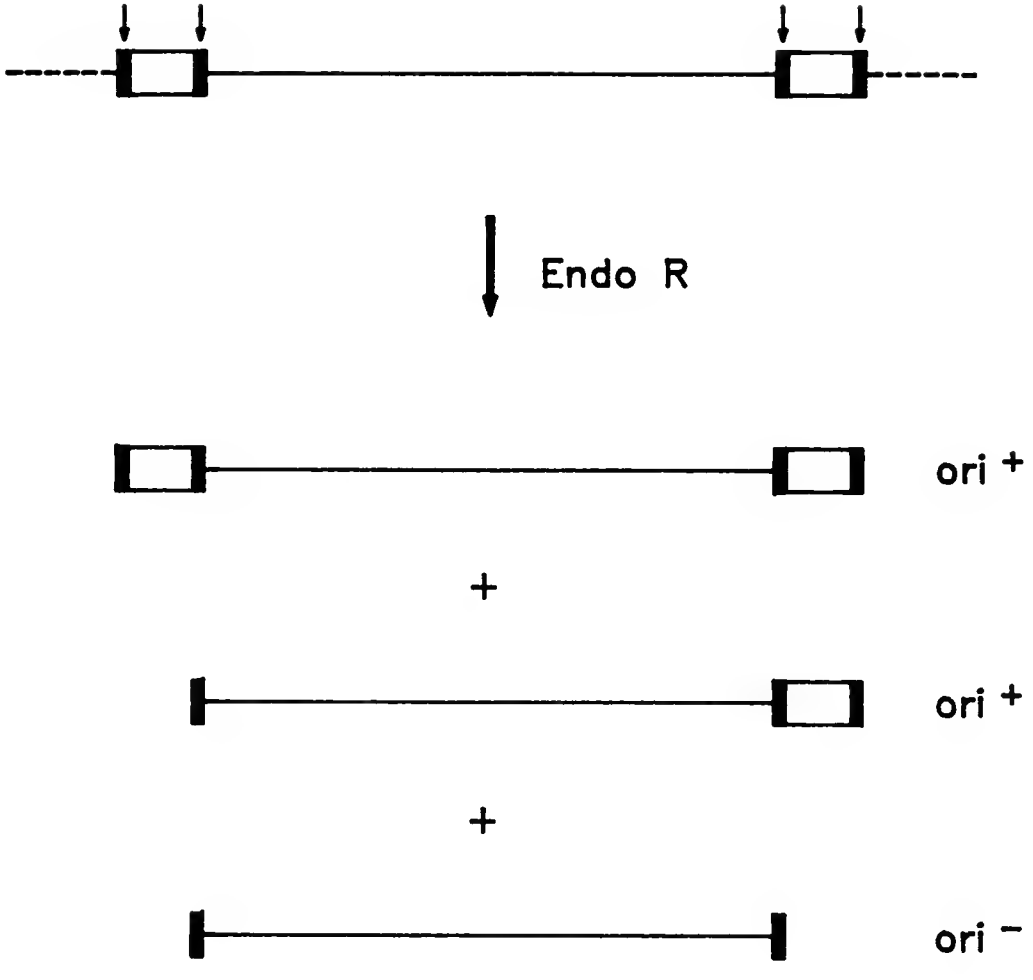
When AAV plasmids are transfected into human cells most of the input plasmid DNA is converted to relaxed circular and linear plasmid DNA. This is true regardless of whether the plasmid contains an AAV genome capable of autonomous replication (Hermonat et al., 1984; Samulski et al., 1982, 1983; Laughlin et al., 1983) or whether the transfection is done in the presence of hydroxyurea. Digestion of the recovered input DNA with restriction enzymes indicates that a large fraction of the input DNA has been cleaved at one or the other AAV junction with vector DNA. Additional fragments seen in vivo suggest that specific cleavage also occurs at the inboard copies of the AAV terminus (the sequence CCaCTCCCTCTCT), and at an internal sequence to produce fragments of 1.6 and 2.4 kb. This means that specific AAV sequences are substrates for a host-encoded endonuclease activity and that this activity does not require DNA replication. All of these in vivo observations are consistent with the properties of endo R in vitro. In addition, cleavage by endo R occurs in a region in AAV substrates that is consistent with what is predicted by the model for AAV DNA replication. This is supported by the observation that endo R products

were substrates for DNA replication *in vivo* and *in vitro*. However, without the benefit of a cellular mutation in endo R, it can not be concluded that this activity is required for AAV rescue.

If endo R represents a major pathway for AAV rescue, then excision of AAV sequences by endo R must explain not only how wild type AAV sequences are rescued, but also how mutants which contain terminal deletions are rescued. In view of the sequence specificity of endo R, the following model can be offered for rescue of wild type and mutant sequences (Figure 7-1). Normal AAV rescue occurs by endo R cleavage at the two outboard AAV recognition sites to directly produce an AAV replicative intermediate. This can occur when the wild type plasmids, pSM620 or pAV1, are transfected into mammalian cells. Cleavage could be directed toward the outboard sites by tandem copies of the AAV terminal sequences between integrated provirus or if AAV integrates into the host genome in a region rich in G:C base pairs. In some molecules, cleavage can occur at an outboard site in one end, and only at an inboard site at the other end (pSM621, pSM609). These molecules can generate an AAV replicative intermediate by subsequent correction of the missing ori sequence (Samulski *et al*., 1983). However, molecules that have been cleaved at both inboard endo R sites are ori^- and incapable of DNA replication. Such molecules would have to recombine with uncleaved input plasmid molecules to recover a functional origin sequence. A prediction of this model is that the deletion of the terminal 15 bp of AAV from both ends (ie., deletion of both outboard recognition sites) should significantly reduce the infectivity of AAV plasmids. This is, in fact, consistent with the behavior of an AAV mutant plasmid ($psub201^+$), recently reported by Samulski *et al*. (1987), in which the

Figure 7-1. Model for the Excision of AAV DNA from Recombinant Plasmids.

Solid line indicates internal AAV sequences, dotted line indicates flanking plasmid sequences, open boxes indicate the AAV origins which consist of the terminal 125 bp palindromic repeats. Vertical lines and arrows indicate the inboard and outboard AAV recognition sequences and endo R cleavage sites.


Endo R
ori +
+
ori +
+
ori −

first 13 bp from are missing both ends. With this plasmid, DNA replication is 8-fold more efficient if the AAV sequences are cleaved away from the vector sequences before transfection into Ad2-infected cells.

The frequency of endo R excision of AAV plasmids is also consistent with what is seen *in vivo*. Based on the observed frequency of cleavage *in vitro*, it would be expected that only 0.01-1% of the input pSM620 molecules would be cut at both junctions to produce ori^+ AAV duplex DNA. In a typical transfection experiment (0.1-10 μg of plasmid DNA per 10 cm dish) an estimated excision frequency of 0.01% would produce 0.3-30 AAV duplex DNA molecules in each transfected cell. (See Chapter II for the calculation.) This is approximately the limit of detection in most transfection experiments. However, if this frequency is correct, then there should be a linear relationship between the yield of replicative form DNA produced in transfected cells and the amount of input plasmid DNA used in the transfection. In agreement with this, Samulski *et al*. (1982) have shown that the yield of AAV DNA is linear with the amount of input DNA in the range between 0.1 and 10 μg of plasmid DNA per 10 cm dish.

In using the recombinant plasmids as substrates for *in vitro* excision, it was assumed that these plasmids were models for AAV proviruses integrated into chromatin. Typically, a cell line carrying AAV proviruses contains 2-5 tandem copies of the AAV genome (Cheung *et al*., 1980; Laughlin *et al*., 1986; McLaughlin *et al*., 1988). The junction of two tandem AAV copies would contain either a single copy of the sequence ggCCaCTCCCTCTCT or two copies of the same sequence in an inverted orientation, producing the palindromic sequence

agagagggagTggCC/ggCCaCTCCCTCTCT. Plasmids containing a single (pGM1228) or a double inverted copy (pGM1344) of this sequence were cleaved by endo R, and the clone containing two copies of the AAV recognition signal was cleaved approximately 5 fold more efficiently. This sequence, which is actually a stretch of polypurine residues followed by a polypyrimidine stretch, is also a derivative of the sequence at the junction of pBR322 and AAV in pSM620 (g_n/CCaCTCCCTCTCT) and this substrate is cleaved at high frequency. This type of recognition sequence could be generated if AAV integrated into a G:C rich region in the host DNA.

Qualitatively, however, there must be at least two major differences between rescue from chromatin and rescue from recombinant plasmids. First, whereas approximately 10^3-10^4 molecules of AAV plasmid DNA are required per cell to see maximum production of AAV virions in a transfection experiment, endogenous AAV proviruses are usually present at less than 10 copies per cell. There appears to be, then, at least 2 orders of magnitude difference between excision frequencies from chromatin and naked DNA. This may reflect the presence of additional components required for excision from chromatin. In this regard, it is worth noting that the efficiency of cleavage at the right AAV/vector junction in pAV1 was significantly better in crude extracts (not shown) than with purified enzyme preparations (Figure 4-3). One possible explanation for this observation is that accessory proteins may have been lost during the purification.

A second issue is the apparent stability of AAV proviruses in the absence of helper virus (Cheung et al., 1980). Although extrachromosomal copies of AAV are occasionally generated during passage

of latently infected cells, most proviral copies are stable indefinitely, unless the cell is superinfected with helper virus (Cheung et al., 1980; McLaughlin et al., 1988). Thus, AAV proviral termini are not usually a target for endo R. Possible explanations for this are that either the termini are inaccessible due to the secondary structure of chromatin or that endo R activity is tightly regulated in the cell. In this regard, Larsen and Weintraub (1982) have shown that the S1 sensitivity of homopurine-homopyrimidine sequences upstream of the chicken beta A globin gene in chromatin depends on whether or not the chromatin was isolated from cells that were actively expressing the globin gene. Thus, the initiation of AAV rescue by adenovirus superinfection may be due to the activation of AAV gene expression by adenovirus gene products (Laughlin et al., 1982; Richardson and Westphal, 1984).

Enzyme Recognition

In general, the level of cutting by endo R depends on the length of the recognition sequence and the amount of substitutions in the G (or C) strand in a given length. Plasmids that contain homopolymeric stretches of poly(dG):poly(dC) are the ideal substrates for endo R. Endo R cleavage experiments using clones that contained stretches of poly(dG) of varying length indicated that a minimum of 9 bases of G residues were required for cleavage and that the frequency of cleavage increased in direct proportion to the length of the homopolymeric chain, to a maximum of approximately 30 base pairs. Substitutions or variations in the ideal sequence reduced the level of cleavage, where the overall frequency depended on the length of the recognition sequence

and the amount of substitution (Table IV). For example, substrates that contained the alternating copolymer poly$(GA)_{38}$ (50% G:C base pairs) were cleaved at 10% of the level observed in substrates containing 31 bp of poly(dG):poly(dC). This reflected the effect of substitution of G residues with A residues. In addition, the level of cleavage seen with plasmids containing poly$(GA)_{38}$ was more than 3 fold greater than the level of cleavage observed with substrates that contained 11 repeats of the same alternating copolymer (poly$(GA)_{11}$), indicating that the length of the substituted polymer also affected the level of cleavage. Moreover, the homopolymer, poly$(dA)_{83}$:poly$(dT)_{83}$, was not a substrate for endo R (not shown). These results suggested that the frequency of cleavage depended on the length of the recognition signal and the percentage of G:C base pairs in this sequence.

The type of substitution in the G:C polymer was also important. Substitution of G residues with A was tolerated, but resulted in significantly lower cleavage frequencies. However, substitutions of G with C residues in the alternating copolymer poly$(GC)_{20}$, completely inhibited double-stranded cleavage (pGM1635, Table IV). On the surface this may seem to indicate that a transition between purines and between pyrimidines was allowed, while substitutions of pyrimidines for purines and purines for pyrimidines was not. However, the comparison of the level of cleavage between clones that contained mixed purine and pyrmidine recognition signals with those that were homopurinic indicated that this was not the case. Substrates that contained the Dictyostelium telomeric sequence, $(C_{2-6}T)_5$, which is a polypyrimidine/polypurine chain (80% C), or the Tetrahymena telomeric sequence, $(C_4A_2)_3$, in which purines are mixed with pyrimidines (67% C), were cleaved at

approximately the same frequency (Table IV), corresponding to about 30% of the level observed with the parental poly(dG)$_{31}$ clone. In addition, the AAV recognition sequence (CCaCTCCCTCTCT), which contains 8 C residues in a 13 base pyrimidine-rich sequence (62% C), was cleaved at 1% of the level observed with the poly(dG)$_{31}$ clone. Thus, the length of the recognition sequence and the degree of deviation from the ideal poly(dG):poly(dC) sequence were major considerations for the level of cleavage activity. However, the type of base substitution was also important in certain cases. This was especially true in cleavage assays with the alternating copolymer (GC)$_{20}$, which can exist in a left handed Z-DNA configuration (Peck and Wang, 1983; Singleton et al., 1983), and was not cleaved by endo R. These results suggested that endo R cleavage frequencies could be better explained in the context of the structural properties of the recognition signal, rather than strictly on a sequence basis.

Evidence from nuclease digestion experiments (Drew, 1984; Cantor and Efstradiatis, 1984; Drew and Travers, 1984; Pulleyblank et al., 1985; Evans et al., 1986) indicate that duplex chains of poly(dG):poly(dC) exhibit a non-B, non-Z DNA conformation, which is similar, but not identical to A-type DNA (Arnott et al., 1974). The conformation of these polymers in solution are characterized by a wide and shallow minor groove and a deep and cavernous hydrated major groove (McCall et al., 1985). The longer helical repeat of approximately 11 bases and the larger helical radius are more typical of double-stranded RNA than B-DNA (Bram, 1971). Under torsional stress or when flanked by B-DNA, these sequences are likely to exist in a ladder-like structure resulting from the stacking of the guanosine bases, while the C-strand

passively swivels around the stacked G-strand (McCall et al., 1985). In addition, the 3′ flanking region of the poly(dG) strand seemed to exist in an "unpaired" conformation in torsionally stress molecules (Kohwi-Shigematsu et al., 1985).

The cleavage behavior of endo R with the various oligomer substrates can be better explained in the context of these structural observations. These observations suggest that the importance of the sequence at the site of recognition may be secondary to that of DNA conformation. This would explain why homopolymers of poly(dG) are the ideal substrates and why certain variations of this sequence are tolerated while others are not. Variations in the recognition sequence result in the disruption of guanosine stacking on the G-strand to different degrees. Duplex molecules containing the alternating copolymer (GA) maintain the unusual stacking conformation of homopolymeric G:C chains to a lessor degree, through purine-purine stacking (Evans et al., 1986). In contrast, in substrates containing alternating GC residues the secondary structure characteristic of poly(G) seems to have been disrupted by the C-residues, and under certain salt conditions, the GC copolymer may exist in a left-handed Z-DNA conformation (Peck and Wang, 1983; Singleton et al., 1983) that differs markedly from the structure of poly(dG):poly(dC). These observations indicate that the mixing of purine and pyrimidine residues in the recognition sequence is not as important as the maintenance of the poly(dG) conformation. This would explain why the telomeric sequences, $(C_4A_2)_3$ and $(C_{2-6}T)_5$, were cleaved at approximately the same frequency. Both sequences presumably maintain enough of the poly(dG) structure with long, interrupted stretches of G:C base pairs, and

neither contain long stretches of alternating GC residues which could disrupt this conformation. In summary, endo R seems to recognize an alternate DNA structure characteristic of homopolymers containing poly(dG):poly(dC). The length and variation of this sequence is important only in the context of how it affects the DNA secondary structure. The substitution of G with A or T residues presumably results in the gradual disruption of this altered conformation, while the introduction of alternating GC residues changes the DNA secondary structure to a form not recognized by endo R.

The analysis of endo R cleavage sites at the sequence level supported the view that enzyme recognition was the result of secondary structure in the DNA molecule. While the frequency of cleavage was shown to be directly proportional to the length and G:C base pair content of the recognition signal, the cleavage sites were distributed, for the most part, in a pattern that was more consistent with structural differences between the various inserts than with the insert sequence. With substrates that contained a high percentage of G:C base pairs (pGM1008, pGM913, pGM1116, pGM1505, Figure 5-6), cleavage occurred between every base of the insert, presumably because minor substitutions of G residues did not cause a significant alteration of the secondary structure. In more highly substituted inserts (pGM1483, pGA38, pGA11, pGM1228), the cleavage pattern apparently reflected alterations of the G:C conformation, resulting in a more specific or periodic array of cleavage sites. For example, cleavage of the sequence $(C_4A_2)_3$ occurred at regular intervals on both strands, immediately 5′ to the T residues on the GT strand (pGM1483, Figure 5-4C, Figure 5-6). This suggested that the poly(dG) structure was locally disrupted by the insertion of

two T residues, while the overall conformation of the insert was unchanged. Therefore, endo R seemed to have the ability to recognize the secondary structure of the entire insert in spite of local disruptions, since none of the uninterrupted stretches of poly(dG) in pGM1483 were long enough to be endo R recognition sites by themselves. This effect was even more pronounced in the alternating poly(GA) clones (pGA38, pGA11, Figure 5-4B, Figure 5-6), where cleavage occurred at every other base in the recognition sequence, 3′ to every A residue (and C residue). This cleavage pattern presumably reflected local perturbations in guanosine base stacking, while purine base interactions maintained an alternate DNA structure in the insert that was still recognized by endo R. The ability of endo R to recognize the secondary structure as a whole, may explain why the stronger cleavage sites in the GA strand of pGA11 and pGA38 (Figure 5-6) occurred at the ends of the insert. The cleavage pattern observed in pGM1228, which contains a single copy of the AAV recognition sequence (Figure 5-4D, Figure 5-6), can also be explained in terms of the secondary structure of the insert. In this clone, cleavage occurred at a low frequency and was confined to a small region, rich in G residues. The short length and relatively high substitution of G residues in this region apparently produced a secondary structure that was only marginally recognized by endo R.

Sequence analysis of the endo R cut sites indicated that cleavage occurred throughout the regions rich in G:C base pairs. However, the pattern of cutting in the G-rich or C-rich strands was not identical and suggested that endo R cleavage of the substrate may occur through a series of concerted single-stranded nicks on opposite strands, rather than by double-stranded cleavage. This was supported by the results

from endo R time course experiments which indicated that the specifically cleaved fragments were the product of double-stranded breaks, while site specific single-stranded nicks in the substrates were not detected. Taken together, these observations imply that the enzyme rapidly produces single-stranded breaks on both strands of the molecule which eventually result in a double-stranded breaks.

The pattern of cleavage in several of the clones (pGM1008, pGM913, pGM1116, pGM1505, pGM1228, and pGM1344, Figure 5-6) depicted a distribution that was skewed toward the 3′ end of each strand (Figure 5-4). This suggested that a significant amount of molecules might contain sizable 3′ protruding ends consisting of C-rich and G-rich sequences. Overhangs of this nature would tend to hybridize together, and in effect, reduce the amount of observable specifically cleaved products in the endo R reactions. The experimental data, however, did not support this conclusion. Phenol extraction and heating of the reaction products before electrophoresis did not increase the yield of specifically cleaved product. However, intramolecular aggregation and basepairing between guanine residues in single-stranded overhangs has been shown to be a thermodynamic possibility (Henderson, et al., 1987). It is possible, then, that basepairing between G residues on one of the 3′ overhangs form hairpin structures that would prevent hybridization of the G-rich and C-rich overhangs.

It was also possible that the products of endo R cleavage contained blunt ends or short 3′ overhangs that were not likely to hybridize. By this mechanism, endo R would produce a double-stranded break in each molecule by making single-stranded nicks at opposing positions in each strand. The distribution of sites throughout the

recognition sequence suggested that very few nicks were made in each strand before a double-stranded break occurred. The cleavage pattern observed in the clones containing the telomeric sequence $(C_4A_2)_3$ (pGM1483, Figure 5-4C, Figure 5-6) and the alternating copolymer $(GA)_{11}$ (pGA11, Figure 5-4B, Figure 5-6) support this hypothesis. Cleavage in these substrates occurred at regular intervals and at opposing positions in each strand. Continued cleavage or degradation of either strand in a 5′ to 3′ direction would account for the increased distribution of cleavage sites at the 3′ end of the insert junction observed in several of the substrates. Thus, the observed distribution of cleavage sites may reflect a population of molecules that are cleaved a minimum number of times anywhere within the recognition sequence. Double-stranded breaks would then be the result of concerted single-stranded nicks in opposing positions on each strand, producing molecules with short 3′-OH overhangs or blunt ends. The higher frequency of cleavage sites at the 3′ end of each strand may be explained as an artifact of additional cleavage or degradation of the cleaved molecules.

The activity profile of endo R is strikingly similar to the activity of a number of single-stranded DNA endonucleases. This is especially true of S1 nuclease, where it seems clear that the enzyme also recognizes the altered secondary structure in duplex DNA consisting of stretches of polypurine/polypyrimidine (Evans and Efstradiatis, 1986; Pulleyblank et al., 1985; Cantor and Efstradiatis, 1984). However, important differences between these activities indicate that endo R is not a classical single-stranded nuclease. First, whereas endo R will cleave short linearized polypurine/polypyrimidine sequences, S1 requires supercoiling of these substrates. This presumably reflects the need for

a torsionally stressed molecule (Cantor and Efstratiadis, 1984) and may indicate that an altered structure of the polymer flanking regions is necessary for cleavage by S1 (Kohwi-Shigematsu and Kohwi, 1985). Secondly, endo R specificity for single-stranded DNA could not be demonstrated by any classical means, including acid solubilization of ^{3}H-*E.coli* DNA or by competition assays with single-stranded DNA. Endo R cleavage of ϕX single-stranded circles may be attributed to an intrinsic activity or may be due to the presence of a contaminating nuclease.

Evidence for a Contaminating Nuclease

Several lines of evidence suggest that endo R preparations contain a contaminating nuclease or a second activity, possibly as an intrinsic part of the same enzyme. For example, background smearing in the standard gel assay indicated that the substrate was being nonspecifically degraded. This effect was observed with even the purest preparations of endo R and became pronounced when large amounts of enzyme were added or with prolonged incubation times. In addition, a significant amount the specifically cleaved linear products were not substrates for T4 DNA ligase and could be 5′ end labeled without prior treatment with CIAP. This indicated that two types of ends were being produced on endo R products. Moreover, examination of endo R products early in the reaction revealed that the form II molecules produced were randomly nicked. A similar random nicking activity was observed on single-stranded ϕX circles. These observations may indicate that a separate, unrelated nucleolytic activity was present in endo R extracts. Alternatively, all of these results can be explained by an alternate

specificity of endo R, rather than the presence of a contaminating nucleolytic activity.

An alternate specificity was demonstrated for endo R under reaction conditions where Mn^{2+} was substituted for Mg^{2+} as metal cofactor. Under these conditions with standard amounts of enzyme, the substrate was degraded non-specifically (Figure 6-3). It is possible that the nonspecific activity coexists, at a low level, with the specific activity under reaction conditions that favor double-stranded specific cleavage. This would explain the nonspecific degradation of the substrate, as well as the random nicking of form II plasmid DNA and single-stranded circular DNA.

A second issue is the inability of a portion of the endo R reaction products to be ligated, and the fact that approximately 20 percent of the specifically cleaved products can be 5′ end labeled without treatment with alkaline phosphatase. These observations indicate that a portion of the product molecules contain 5′ hydroxyl groups. It is unlikely, however, that this is the result of a separate nucleolytic activity, since both the phosphorylated and dephosphorylated ends occur on specifically cleaved molecules (Figure 6-10B). It is possible that dephosphorylation of the specifically cleaved molecules is due to an associated or contaminating phosphatase activity or that a small amount of exonucleolytic activity which leaves 3′ hydroxyl groups is present in the extracts.

Other Mammalian Endonucleases

The characterization of site-specific endonucleases, which recognize specific DNA sequences or minor distortions in the DNA helix,

has mostly been confined to those obtained from procaryotic sources. Recently, however, a number of laboratories have isolated endonuclease from eucaryotic cells that show at least some specificity for G residues (McKenna et al., 1981; Desiderio and Baltimore, 1984; Kataoka et al., 1984; Ruiz-Carillo and Renaud, 1987). In some respects, endo R is different from all of these activities. For example, the non-random endonuclease isolated by McKenna et al. (1981) appears to be a degradative enzyme which digests adenovirus and SV40 DNA and initially produces discrete fragments which are reduced to short oligonucleotides with further digestion. The fragments produced have predominantly G residues at their ends. In contrast, endo R has little if any activity on Ad2 and SV40 DNA. Similarly, an endonucleolytic activity which cleaves near the recombination site of immunoglobulin J_K segments (Kataoka et al., 1984), also cleaves sites in pBR322 that consist of short poly(dG) stretches (4-7 residues), while endo R does not detectably cleave these short poly(dG) sequences. The activity most similar to endo R is the one recently described by Ruiz-Carillo and Renaud (1987) which they have called endo G. Endo G was isolated from chicken erythrocytes and appears to cleave exclusively at stretches of poly(dG)-poly(dC) which have a minimum length of 10 bp. Unlike endo R, endo G does not appear to cleave alternating homopurine/homopyrimidine stretches (poly(ga)-poly(CT). In spite of these differences, it is possible that endo R is related to some and perhaps all of these previously described enzymes. None of these enzymes have been purified to homogeneity and their properties may change significantly during the course of purification. By analogy with the recBC endonuclease (Rosamond et al., 1979; Taylor and Smith, 1980; Muskavitch and Linn,

1981; Taylor et al., 1985), it is possible that all of these endonucleases share a common core activity whose specificity is modified by inhibitory or accessory proteins or the conditions of the reaction.

What Does Endo R Do for the Cell

The isolation of an enzyme from HeLa cells whose sole activity appears to be cleavage at homopurine/homopyrimidine sequences raises the question of the role of these sequences in cellular DNA. Endo R is one of three major endonucleolytic activities that were detected in HeLa cell nuclei. However, because of the specificity of endo R, it is unlikely that it plays a degradative role in the cell. Furthermore, the results of Larsen and Weintraub (1982) suggest that most poly(dG) or poly(GA) sequences in chromatin would be protected from cleavage unless they are undergoing transcription or replication.

The distribution of homopurine/homopyrimidine sequences appears to be predominantly in non-coding sequences (Nakamura et al., 1987; Hoffman-Lieberman et al., 1986; Jeffreys et al, 1985;), but their distribution does not suggest any obvious role. In some cases these sequences have been found near coding regions (Hentschel, 1982; Gliken et al., 1983; Richards et al., 1983; Mace et al., 1983; Field et al., 1984; Gerondakis et al., 1984; Shih et al., 1984; McGhee et al., 1981; Day et al., 1981). An intriguing aspect of these sequences is that in many cases they have been found in regions of genetic instability, either as part of satellite DNA (Fowler and Skinner, 1986) or as part of moderately repeated tandem arrays which vary in the number of repeat units (Jeffreys et al., 1985; Nakamura et al., 1987; Hoffman-Liebermann et al., 1986). Some of the variable number tandem repeat units (VNTR's)

have been found near coding genes as in the case of the insulin gene (Bell et al., 1982) and the myoglobin gene (Jeffreys et al., 1985).

One commonly proposed mechanism for amplifying or reducing the number of repeats in a tandem array is unequal homologous recombination. This has led to the suggestion that VNTR sequences may contain hotspots for recombination (Jeffreys et al., 1985). Although there is no direct evidence to support this, it is striking that many of the VNTR's have homopurine/homopyrimidine stretches similar to those found in AAV. Jeffreys et al. (1985) concluded that the common core sequence in their repeats was CxTCCTgCCC. Nakamura et al. (1987) deduced a somewhat different core sequence, CCCCaCnnCCC. Specific examples of sequences that have been associated with variable numbers of tandem repeats are the insulin gene, CCCCaCaCCCC (Bell et al., 1982), the myoglobin gene, CCTCCaCCCgTCCTT (Jeffreys et al., 1985), and the DR2 repeat of the herpes simplex 1 terminal sequence, CTCCTCCCCCC (Mokarski and Roizman, 1983). It is possible, then, that endo R recognizes the secondary structure of the VNTR sequences and is involved in one of the cellular pathways for recombination. If, in fact, AAV termini contain a recognition sequence for cellular recombination, then this might explain how AAV proviruses are both integrated and rescued.

The sequence recognition of endo R is reminiscent of the properties of telomere terminal transferase (telomerase) from Tetrahymena (Henderson et al., 1987). Although some variation exists in the consensus sequence, all known nuclear telomeres consist of G-rich and C-rich complementary strands (Blackburn, 1984). Several variations of this sequence are recognized as telomeres in yeast, suggesting that all of these sequences maintain a common secondary structure recognized

by telomerase (Szostak and Blackburn, 1982; Murray and Szostak, 1983; Shampay et al., 1984; Blackburn and Szostak, 1984). In addition, elongation by telomerase requires telomeric sequences, at least 12 bp in length, that contain 3' protruding ends rich in G-residues (Blackburn, 1986; Henderson et al., 1987).

It is conceivable that endo R also plays a role in the maintenance of chromosomal ends. Two variations of telomeric sequences were substrates for endo R cleavage (pGM1483 and pGM1505, Figure 5-4C, Figure 5-6), and it is likely that all the variations of these sequences would also be recognized and cleaved by endo R as well. In addition, the sequence analysis of endo R cleavage sites, suggested that the enzyme left 3' protruding ends. A portion of these would be rich in G residues, and would, therefore, be substrates for the nontemplated addition of sequences by telomerase. Thus, endo R cleavage of telomeric sequences during DNA replication could allow the resolution of the ends of the replicated chromosome, and at the same time, prepare telomeric sequences for strand elongation by telomerase.

The results from studies of the effect of nucleotides and polynucleotides on endo R activity suggest that endo R cleavage in vivo may be tightly regulated. Endo R displayed a strong affinity for homopolymers of single-stranded poly(dG) in DNA competition experiments. This supported the conclusion that the major substrate for endo R was poly(dG). However, the inhibition by poly(dG) may also reflect autoregulatory properties of the enzyme. The accumulation of small oligonucleotides of poly(dG) produced from endo R cleavage could serve to inhibit or negatively modulate the cleavage reaction. Thus, the amount of cleavage by endo R may be limited by endproduct inhibition.

Additional evidence suggests that endo R activity may be allosterically modulated by ribo- and deoxyribo-nucleotides. Cleavage activity was stimulated 2-3 fold with the addition of dGTP and dATP, while the activity was strongly inhibited with the addition of ATP and GTP. However, endo R did not bind to ATP- or GTP-linked agarose columns and it is possible that the binding of these compounds by endo R requires the presence of substrate DNA. These observations suggest that there are two separate binding sites on the enzyme. One site may recognize and bind DNA secondary structure, while another site may bind substances that effect cleavage activity indirectly, possibly by causing conformational changes of the enzyme. This level of control may be necessary to prevent the cleavage of homopurine/homopyrimidine sequences near the coding regions of transcriptionally active genes, when the concentration of ribonucleotides is relatively high and the sequences flanking the coding regions are exposed (Larsen and Weintraub, 1982). During cellular DNA replication however, higher local concentrations of deoxyribonucleotides may stimulate endo R cleavage activity.

In conclusion, endo R is a human cellular enzyme most likely consisting of a single peptide with a native molecular weight of approximately 120 kd. The enzyme apparently recognizes the secondary structure characteristic of polypurine/polypyrimidine sequences rich in G:C base pairs. Sequence analysis of the cleavage site suggests that endo R produces double-stranded breaks in the substrate through a series of concerted single-stranded nicks occurring throughout the recognition site. The properties of the enzyme and the distribution of naturally

occurring recognition sequences in chromosomal DNA, suggest that Endo R cleavage activity may play a role in genetic recombination and cellular DNA replication.

REFERENCES

Arnott, S., Chandrasekaran, R., Hall, I.H., and Puigjaner, L.C. (1983) Heteronomous DNA. Nucl. Acids Res. 11, 4141-4155.

Atchison, R.W., Casto, B.C., and Hammond, W.McD. (1965) Adenovirus-associated defective virus particles. Science 149, 754-756.

Bell, G.I., Selby, M.J., and Rutter, W.J. (1982) The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. Nature 295, 31-35.

Berns, K.I., Muzyczka, N., and Hauswirth, W.W. (1985) Replication of parvoviruses, in Virology, Bernard Fields, ed., Raven Press, New York, pp.415-432.

Berns, K.I., Pinkerton, T.C., Thomas, G.F., and Hoggan, M.D. (1975) Detection of adeno-associated virus (AAV)-specific nucleotide sequences in DNA isolated from latently infected Detroit 6 cells. Virology 68, 556-560.

Blackburn, E.H. (1984) Telomeres: Do the ends justify the means? Cell 37, 7-8.

Blackburn, E.H. (1986) Molecular structure of telomeres in lower eukaryotes, in Molecular Developmental Biology, L. Bogorad, ed., New York: Alan R. Liss, Inc., pp.69-82.

Blackburn, E. H. and Szostak, J.W. (1984) The molecular structure of centromeres and telomeres. Annu. Rev. Biochem. 53: 163-194.

Boissy, R. and Astell, C.R. (1985) *Escherichia coli recBC sbcB recF* hosts permit the deletion-resistant propagation of plasmid clones containing the 5'-terminal palindrome of minute virus of mice. Gene 35, 179-185.

Bolivar, F., Rodriguez, R.L., Greene, D.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H., and Falkow, S. (1979) Construction and characterization of new cloning vectors. II. A multipurpose cloning system. Gene 2, 95.

Boyer, H.W. (1969) A complementation analysis of the restriction and modification of DNA in Eschericia Coli. J. Mol. Biol. 41, 459-479.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248-254.

Bram, S. (1971) Polynucleotide polymorphism in solution. Nature New Biol. 233, 161-164.

Buller, R.M., Janik, J.E., Sebring, E.D., and Rose, J.A. (1981) Herpes simplex virus types 1 and 2 completely help adenovirus-associated virus replication. J. Virol. 40, 241-247.

Cantor, C.R. and Efstradiatis, A. (1984) Possible Structures of homopurine-homopyrimidine S1-hypersensitive sites. Nuc. Acids Res. 12, 8059-8072.

Cavalier-Smith, T. (1974) Palindromic base sequences and replication of eukaryotic chromosome ends. Nature 250, 467-470.

Challberg, M.D. and Kelly, T.J. (1979) Adenovirus DNA replication *in vitro*. Proc. Natl. Acad. Sci. U.S.A. 76, 655-659.

Chen, G.L., Yang, L., Rowe, T.C., Halligan, B.D., Tewey, K.M., and Liu, L.F. (1984) Nonintercalative antitumor drugs interfere with the breakage-reunion reaction of mammalian DNA topoisomerase II. J. Biol. Chem. 259, 13560-13566.

Cheung, A.K., Hoggan, M.D., Hauswirth, W.W., and Berns,K.I. (1980) Integration of the adeno-associated virus genome into cellular DNA in latently infected human Detroit 6 cells. J. Virol. 33, 739-748.

Day, L.E., Hirst, A.J., Lai, E.C., Mace, M., and Woo, S.L.C. (1981) 5′ domain and nucleotide sequence of an adult chicken chromosomal B-globin gene. Biochemistry 20, 2091-2098.

Desiderio, S. and Baltimore, D. (1984) Double-stranded cleavage by cell extracts near recombinational signal sequences of immunoglobulin genes. Nature 308, 860-862.

Drew, H.R. (1984) Structural specificities of five commonly used DNA nucleases. J. Mol. Biol. 176, 535-557.

Drew, H.R. and Travers, A.A. (1984) DNA structural variations in the *E.coli tyr*T promoter. Cell 37, 491-502.

Evans, T. and Efstradiatis, A. (1986) Sequence-dependent S1 nuclease hypersensitivity of a heteronomous duplex. J. Biol. Chem. 261, 14771-14780.

Field, L.J., Philbrick, W.M., Howles, P.N., Dickinson, D.P., McGowen, R.A., and Gross, K.W. (1984) Expression of tissue-specific Ren-1 and Ren-2 genes of mice: Comparative analysis of 5′-proximal flanking regions. Mol. Cell. Biol. 4, 2321-2331.

Fowler, R.F. and Skinner, D.M. (1986) Eucaryotic DNA diverges at a long and complex pyrimidine-purine tract that can adopt altered conformations. J. Biol. Chem. 261, 8994-9001.

Gerondakis, S., Cory, S., and Adams J.M. (1984) Translocation of the myc cellular oncogene to the immunoglobulin heavy chain locus in murine plasmacytomas is an imprecise reciprocal exchange. Cell 36, 973-982.

Glikin, G.C., Gargiulo, G., Rena-Descalzi, L., and Worcel, A. (1983) Escherichia coli single-strand binding protein stabilizes specific denatured sites in superhelical DNA. Nature 303, 770-774.

Green, M., Parsons, J.T., Pina, M., Fujinagu, K., Caffin, A., and Landgraf-Lears, M. (1971) Transcription of adenovirus genes in productively infected and transformed cells. Cold Spring Harbor Symp. Quant. Biol. 35, 803.

Handa, H., Shiroki, K., and Shimojo, H. (1977) Establishment and characterization of KB cell lines latently infected with adeno-associated virus type 1. Virology 82, 84-92.

Hauswirth, W.W. and Berns, K.I. (1977) Origin and termination of adeno-associated virus DNA replication. Virology 78, 488-99.

Hauswirth, W.W., van de Walle, M.J., Laipis, P.J., and Olivo, P.D. (1984) Heterogeneous mitochondrial DNA D-loop sequences in bovine tissue. Cell 37, 1001-1007.

Henderson, E., Hardin, C.C., Walk, S.K., Tinoco Jr., I., and Blackburn, E.H. (1987) Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. Cell 51, 899-908.

Hentschel, C.C. (1982) Homocopolymer sequences in the spacer of a sea urchin histone gene repeat are sensitive to S1 nuclease. Nature 295, 714-716.

Hermonat, P.L. and Muzyczka, N. (1984) Use of adeno-associated virus as a mammalian DNA cloning vector: Transduction of neomycin resistance into mammalian tissue culture cells. Proc. Natl. Acad. Sci. USA 81, 6466-6470.

Hermonat, P.L., Labow, M.A., Wright, R., Berns, K.I., and Muzyczka, N. (1984) Genetics of adeno-associated virus: isolation and preliminary characterization of adeno-associated virus type 2 mutants. J.Virol. 51, 329-339.

Hirt, B. (1967) Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. 26, 365-369.

Hoffman-Liebermann, B., Hoffman-Liebermann, D., Troutt, A., Kedes, L.H., and Cohen,S.N. (1986) Human homologs of TU transposon Sequences: Polypurine/Polypyrimidine sequence elements that can alter DNA conformation in vitro and in vivo. Mol. Cell. Biol. 6, 3632-3642.

Hoggan, M.D., Thomas, G.F., Thomas, F.B., and Johnson, F.B. (1972) Continuous carriage of adenovirus associated virus genome in cell culture in the absence of helper adeno-virus. In Proceedings of the fourth lepetite colloquium. Cocoyac, Mexico. North-Holland Publishing Company, Amsterdam, p.243-249

Holloman, W.K., Rowe, T.C., and Rusche, J.R. (1981) Studies on Nuclease alpha from Ustilago maydis. J. Biol. Chem. 256, 5087-5094.

Jeffreys, A.J., Wilson, V., and Thein, S.L. (1985) Hypervariable "minisatellite" regions in human DNA. Nature 314, 67.

Kataoka, T., Kondo,S., Nishi, M., Kodaira, M., and Honjo, T. (1984) Isolation and characterization of endonuclease J: a sequence-specific endonuclease cleaving immunoglobulin genes. Nuc. Acids Res. 12, 5995-6010.

Kohwi-Shigematsu, T. and Kohwi, Y. (1985) Poly(dG)-Poly(dC) sequences, under torsional stress, induce an altered DNA conformation upon neighboring DNA sequences. Cell 43, 199-206.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227, 680-685.

Larsen, E. and Weintraub, H. (1982) An altered DNA conformation detected by S1 Nuclease occurs at specific regions in active chick globin chromatin. Cell 29, 609-622.

Laskowski, Sr., M. (1985) Nucleases: Historical Perspectives, in Nucleases, Linn, M.L. and Roberts, R.J., ed., Cold Spring Harbor Laboratory, pp.2-21.

Laughlin, C.A., Cardellichio, C.B., and Coon, H.C. (1986) Latent infection of KB cells with adeno-associated virus type 2. J. Virol. 60, 515-524.

Laughlin, C.A., Jones, N., and Carter, B.J. (1982) Effect of deletions in adenovirus region 1 genes upon replication of adeno-associated virus. J. Virol. 41, 868-876.

Laughlin, C.A., Tratschin, J.D., Coon, H., and Carter, B.J. (1983) Cloning of infectious adeno-associated virus genomes in bacterial plasmids. Gene 23, 65-73.

Lefebvre, R.B., Riva, S., and Berns, K.I. (1984) Conformation takes precedence over sequence in adeno-associated virus DNA replication. Mol. Cell. Biol. 4, 1416-1419.

Li, J.J. and Kelly, T.J. (1984) Simian virus 40 DNA replication *in vitro*. Proc. Natl. Acad. Sci. USA 81, 6973-6977.

Liu, L.F. and Davis, J.L. (1981) Novel topologically knotted DNA from bacteriophage P4 capsids: Studies with DNA topoisomerases. Nucl. Acids Res. 9, 3979-3989.

Liu, L.F., Liu, C-C., and Alberts, B.M. (1980) Type II DNA topoisomerases: Enzymes that can unknot a topologically knotted DNA molecule via a reversible double-stranded break. Cell 19, 697-707.

Liu, L.F., Rowe, T.C., Yang, L., Tewey, K.M., and Chen, G.L. (1983) Cleavage of DNA by mammalian DNA topoisomerase II. J. Biol. Chem. 258, 15365-15370.

Lusby,E., Fife,K.H., and Berns, K.I. (1980) Nucleotide sequence of the inverted terminal repetition in adeno-associated virus DNA. J. Virol. 34, 402-409.

Mace, H.A.F., Pelham, H.R.B., and Travers, A.A. (1983) Association of an S1 nuclease-sensitive structure with short direct repeats 5'of Drosophila heat shock genes. Nature 303, 555-557.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, New York: Cold Spring Harbor Laboratory).

Maxam, A.M. and Gilbert, W. (1977) A new method of sequencing DNA. Proc. Nat. Acad. Sci. USA 74, 560-564.

McCall, M., Brown, T., and Kennard, O. (1985) The crystal structure of d(G-G-G-G-C-C-C-C): A model for poly(dG):poly(dC). J. Mol. Biol. 183, 385-396.

McCutchan, J.H. and Pagano, J.S. (1968) Enhancement of the infectivity of SV40 DNA with DEAE-dextran. J. Nat. Cancer Inst. 41, 351-357.

McGee, J.D., Wood, W.I., Dolan, M., Engel, J.D., and Felsenfeld, G. (1981) A 200 base pair region at the 5' end of the chicken adult B-globin gene is accessible to nuclease digestion. Cell 27, 45-55.

McKenna, W.G., Maio, J.J., and Brown, F.L. (1981) Purification and properties of a mammalian endonuclease showing site-specific cleavage of DNA. J. Biol. Chem. 256, 6435-6443.

McLaughlin, S.K., Collis, P., Hermonat, P.L., and Muzyczka, N. (1988) Adeno-associated virus general transduction vectors: Analysis of proviral structures. J. Virol. 62, 1963-1973.

McPherson, R.A., Rosenthal, L.J., and Rose, J.A. (1985) Human cytomegalovirus completely helps adeno-associated virus replication. Virology 147, 217-222

Mocarski, E.S. and Roizman, B. (1981) Site-specific inversion sequence of the herpes simplex virus genome: Domain and structural features. Proc. Natl. Acad. Sci. U.S.A. 78, 7047-7051.

Murakami, Y., Eki, T., Yamada, M., Prives, C., and Hurwitz, J. (1986) Species-specific *in vitro* synthesis of DNA containing the polyoma virus origin of replication. Proc. Natl. Acad. Sci. USA **83**, 6347-6351.

Murray, A.W. and Szostak, J.W. (1983) Construction of artificial chromosomes in yeast. Nature **305**, 189-193.

Muskavitch, K.M.T. and Linn, S. (1981) *Rec*BC-like enzymes: Exonuclease V deoxyribonucleases, in The Enzymes, Boyer P.D. ed., Academic Press, New York, p.233.

Muzyczka, N. (1980) Construction of an SV40-derived cloning vector. Gene **11**, 63-77.

Nakamura, Y., Leppert, M., O'Connell, P., Wolfe, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., and White, R. (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. Science **235**: 1616-1622.

Nelson, E.M., Tewey, K.H., and Liu, L.F. (1984) Mechanism of antitumor drug action: Poisoning of mammalian DNA topoisomerase II on DNA by 4'-(9-acridinylamino)-methanesulfon-*m*-anisidide. Proc. Natl. Acad. Sci. USA **81**, 1361-1365.

Peck, L.J. and Wang, J.C. (1983) Energetics of B-to-Z transitions in DNA. Proc. Natl. Acad. Sci. USA **80**, 6206-6210.

Pulleyblank, D.E., Haniford, D.B., and Morgan, A.R. (1985) A structural basis for S1 nuclease sensitivity of double-stranded DNA. Cell **42**, 271-280.

Richards, J.E., Gilliam, A.C., Shen, A., Tucker, P.W., and Blattner, F.R. (1983) Unusual sequences in the murine immunoglobulin u-g heavy-chain region. Nature **306**, 483-487.

Richardson, W.D. and Westphal, H. (1985) Adenovirus early gene regulation and the adeno-associated virus helper effect. Curr. Top. Microbiol. Immunol. **109**, 147-165.

Rosamond, J., Telander, K.M., and Linn, S. (1979) Modulation of the action of the recBC enzyme of *Escherichia coli* K-12 by Ca^{2+}. J. Biol. Chem. **254**, 8646-8652.

Ruiz-Carillo A. and Renaud J. (1987) Endonuclease G: a $(dG)_n:(dC)_n$-specific DNase from higher eukaryotes. EMBO J. **6**, 401-407.

Samulski, R.J., Berns, K.I., Tan. M., and Muzyczka, N. (1982) Cloning of adeno-associated virus into pBR322: rescue of intact virus from the recombinant plasmid in human cells. Proc. Natl. Acad. Sci. U.S.A. **79**, 2077-81.

Samulski, R.J., Chang, L-S, and Shenk, T. (1987) A recombinant plasmid from which an infectious Adeno-Associated Virus genome can be excised *in vitro* and its use to study viral recombination. J. Virol. 61, 3096-3101.

Samulski, R.J., Srivastava, A., Berns, K.I., and Muzyczka, N. (1983) Rescue of adeno-associated virus from recombinant plasmids: Gene correction within the terminal repeats of AAV. Cell 33, 135-143.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Nat. Acad. Sci. USA 74, 5463-5467.

Schon, E., Evans, T., Welsh, J., and Efstradiatis, A. (1983) Conformation of Promoter DNA: fine mapping of S1-hypersensitive sites. Cell 35, 837-848.

Semler, B.L., Dorner, A.J., and Wimmer, E. (1984) Production of infectious poliovirus cloned cDNA is dramatically increased by SV40 transcription and replication signals. Nucl. Acids Res. 12, 5123-5141.

Senapathy, P., Tratschin, J.D., and Carter, B.J. (1984) Replication of adeno-associated virus DNA: Complementation of naturally occurring rep- mutants by a wild-type genome or an ori- mutant and correction of terminal palindrome deletions. J. Mol. Biol. 179, 1-20.

Shampay, J., Szostak, J.W., and Blackburn E.H. (1984) DNA Sequences of telomeres maintained in yeast. Nature 310, 154-157.

Shih, C-K., Linial, M., Goodenow, M.M., and Hayward, W.S. (1984) Nucleotide sequence 5′ of the chicken c-myc coding region: Localization of a non-coding exon that is absent from myc transcripts in most avian leukosis virus-induced lymphomas. Proc. Natl. Acad. Sci. U.S.A. 81, 4697-4701.

Singleton, C.K., Klupik, J., and Wells, R.D. (1983) Conformational flexibility of junctions between contiguous B- and Z-DNA in supercoiled plasmids. Proc. Natl. Acad. Sci. USA 80, 2447-2451.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503-518.

Srivastava, A., Lusby, E.W., and Berns, K.I. (1983) Nucleotide sequence and organization of the adeno-associated virus 2 genome. J. Virol. 45, 555-564.

Straus, S.E., Sebring, E.D., and Rose, J.A. (1976) Concatemers of alternating plus and minus strands are intermediates in adenovirus-associated DNA synthesis. Proc. Natl. Acad. Sci. U.S.A. 73, 742-746.

Switzer, R.C., Muril, C.R., and Shifrin, S. (1979) A highly sensitive silver stain for detecting proteins and peptides in acrylamide gels. Anal. Biochem. 98, 231.

Szostak, J.W. and Blackburn, E.H. (1982) Cloning yeast telomeres on linear plasmid vectors. Cell 29, 245-255.

Taylor, A.F. and Smith, G.R. (1980) Unwinding and rewinding of DNA by the recBC enzyme. Cell 22, 447-457.

Taylor, A.F., Schultz, D.W., Ponticelli, A.S., and Smith, G.R. (1985) RecBC enzyme nicking at chi sites during DNA unwinding: Location and orientation dependence of cutting. Cell 41, 153-163.

Timson, J. (1975) Hydroxyurea (Review). Mutation Research 32, 115-132.

Vogt, V. (1973) Purification and further properties of single-stranded specific nuclease from Aspergillus oryzae. Eur. J. Biochem. 33, 192-200.

Yang. V.C. and Langer, R. (1987) A simple and economical technique for pI measurement. Biotechniques 5, 138-144.

BIOGRAPHICAL SKETCH

Jonathan was born in New York, New York, and was raised on Long Island, New York. He graduated from Manhasset High School, Manhasset, New York, in June 1974 and entered Suffolk County Community College in September of that year. In September of 1976, John transferred to the State University of New York at Albany and received Bachelor of Science degrees in biology and chemistry in August of 1980. From 1980 to 1982, he studied graduate biochemistry in a Master of Science program at the same university. In September of 1982, John moved to Gainesville, Florida, and entered graduate school in the Department of Immunology and Medical Microbiology at the University of Florida. He thought he had left Long Island for good, but in August, 1985, his mentor, Nicholas Muzyczka, moved to the State University of New York at Stony Brook, where John completed his research and earned his Ph.D. from the University of Florida at Gainesville in August, 1988.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Nicholas Muzyczka, Ph.D., Chairman
Associate Professor of Immunology
and Medical Microbiology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Edward K. Wakeland, Ph.D
Associate Professor of Pathology
and Laboratory Medicine

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

William W. Hauswirth, Ph.D
Professor of Immunology and
Medical Microbiology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Thomas C. Rowe, Ph.D
Assistant Professor of
Pharmacology and Therapeutics

This dissertation was submitted to the Graduate Faculty of the College of Medicine and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1988

Dean, College of Medicine

Dean, Graduate School

UNIVERSITY OF FLORIDA
3 1262 08554 3956